A Statistical Study of the Heterotic Landscape

Keith R. Dienes
University of Arizona

---

with Michael Lennek
David Senechal
Vaibhav Wasnik

- hep-th/0602286 (PRD)
- hep-th/0610319 (PRD)
- arXiv: 0704.1320
- to appear

UC Davis seminar, 5/21/2007

Outline

- Introduction/motivation/apology
- N=0 models
  - gauge group correlations
  - Λ correlations
  - correlations between gauge groups and Λ
- N=1, N=2, N=4 results
  - correlations with amounts of SUSY
- Statistical challenges in extracting correlations from a partial data set
- Conclusions, prospects, and warnings

String theory gives rise to a multitude of self-consistent string vacua…

- Correspond to different choices of compactification manifolds/orbifolds, Wilson lines, vevs for numerous moduli fields, fluxes, etc.
- Historically, such vacua were not stable, but it was believed that a stabilization mechanism and/or vacuum selection principle would be found.
- Recent developments suggest that a plethora of vacua continue to exist even after stabilization. KKLT

Such vacua can be viewed as local minima of a complex terrain of hills and valleys …

the string-theory landscape.

The *real* string landscape...

Tucson, Arizona

10^{500} = Pretty darn big

10000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

00000000000000000000000000000000000000000000000000

Does it matter?

Yes!

The low-energy phenomenology that emerges from the string depends critically on the particular choice of vacuum state.

Detailed quantities such as

- choice of gauge group
- number of chiral generations
- SUSY-breaking scale
- cosmological constant, etc.

...all depend on the particular vacuum state selected.

How then can we make progress in the absence of a vacuum selection principle?

Recent proposal: Examine the landscape *statistically*, look for correlations between low-energy phenomenological properties that would otherwise be unrelated in field theory.

Douglas,...

This then provides a new method for extracting phenomenological predictions from string theory.

This idea has triggered a surge of activity examining the statistical properties of the landscape...

- SUSY-breaking scale
- Cosmological constant
- Ranks of gauge groups
- Prevalence of SM gauge group
- Numbers of chiral generations, etc.

Douglas, Dine, Gorbatov, Thomas, Denef, Giryavets, de Wolfe, Kachru, Tripathy, Conlon, Quevedo, Kumar, Wells, Taylor, Acharya, Gorbatov, Blumenhagen, Gmeiner, Honecker, Lust, Weigand, Dijkstra, Huiszoon, Schellekens, Nilles, Raby, Ratz, Wingerter, Faraggi,...

This line of attack has also led to various paradigm shifts...

- Alternative notions of naturalness
- New cosmo/inflationary scenarios
- Anthropic arguments
- Field-theory analogues
- Landsape versus swampland
- Land-skepticism

Douglas, Dine, Gorbatov, Thomas, Weinberg, Susskind, Bousso, Polchinski, Feng, March-Russell, Sethi, Wilczek, Firouzjahi, Sarangi, Tye, Kane, Perry, Zytkow, KRD, Dudas, Gherghetta, Arkani-Hamed, Dimopoulos, Kachru, Freivogel, Vafa, Banks,...

The String Vacuum Project (SVP)

A large, multi-year, multi-institution, interdisciplinary collaboration to explore the space of string vacua, compactifications, and their low-energy implications through

- enumeration and classification of string vacua
- detailed analysis of those vacua with realistic low-energy phenomenologies
- statistical studies across the landscape as a whole.

Will involve intensive research at the intersection of

- *Particle physics:* string theory and string phenomenology
- *Mathematics:* algebraic geometry, classification theory
- *Computer science:* algorithmic studies, parallel computations, database management.

THE STRING VACUUM PROJECT (SVP)

Michael R. Douglas   Gordon Kane

Nima Arkani-Hamed	Mirjam Cvetic	Keith R. Dienes
Michael Dine	Steve Giddings	Shamit Kachru
Paul Langacker	Joe Lykken	Burt Ovrut
Stuart Raby	Lisa Randall	Gary Shiu
Washington Taylor	Henry Tye	Herman Verlinde

September 18, 2006

Abstract

The time is ripe for bringing systematic methods to bear on the construction and analysis of compactifications of string theory as models of realistic particle physics. We propose to pursue a systematic study of the space of string compactifications leading to four-dimensional physics with a series of focused

- **Wiki at:** `http://strings0.rutgers.edu:8000`
- **European SVP website at:**

  `http://www.ippp.dur.ac.uk/~dgrell/svp`

Unfortunately, although there have been many abstract theoretical discussions of string vacua and their statistical properties, there have been very few direct statistical examinations of actual string vacua.

In spite of recent progress, this is because the construction and analysis of actual string vacua remains a fairly complicated affair.

There have been exceptions, however…

- A recent computer analysis of millions of supersymmetric intersecting D-brane models on a particular orientifold background
  - although these models are not stable (they have flat directions), statistical occurrences of various gauge groups, chirality, numbers of generations, etc. were reported.

Blumenhagen, Gmeiner, Honecker, Lust, Weigand

- A similar study focusing on Gepner-type orientifolds exhibiting chiral MSSM spectra

Dijkstra, Huiszoon, Schellekens

Before our work, however, there were almost no studies of the *heterotic* landscape. This is somewhat ironic, since perturbative heterotic strings were the framework in which most of the original work in string phenomenology was performed in the late 1980's and early 1990's.

Moreover, heterotic models are fundamentally different from Type I models...

- tighter constraints (central charges, modular invariance, ...)
- gauge groups generated differently, maximal ranks
- different phenomenologies (e.g., gauge coupling unification)

Expect potentially different statistical properties/correlations.
(May even provide useful guides for heterotic model-builders.)

In this talk, we shall discuss the results of the first statistical study of the heterotic landscape.

- We shall begin by focusing on a sample size of 10^5 distinct four-dimensional non-supersymmetric heterotic string models.
- We shall then enlarge our discussion to include tens of millions of four-dimensional heterotic string models with N=0,1,2,4 SUSY.
- Finally, we shall discuss general statistical issues and problems that affect analyses of this type.

Since our work, other statistical examinations of various portions of the heterotic landscape have also appeared --

- Lebedev, Nilles, Raby, Ramos-Sanchez, Ratz, Vaudrevange, Wingerter
- Faraggi, Kounnas, Rizos

First, three apologies...

- Sample sizes are relatively small, but state of the art
- In this talk, will only concentrate on gauge groups and one-loop vacuum amplitudes (cosmological constants) --- analysis of other features (particle representations, Yukawa couplings, etc.) in progress
- Models not stable, thus not the sort of models we ideally would like to be studying!
  - all N=0 models are tachyon-free, thus stable at tree level, but certainly not stable beyond this
  - even SUSY models have flat directions

On the other hand...

- All models are self-consistent at tree level
  - conformal/modular invariance, proper GSO projections, proper spin-statistics relations, etc.

- Models range from very simple to extraordinarily complex, with many overlapping layers of orbifold twists and Wilson lines, all randomly generated but satisfying tight self-consistency constraints
  - Such high degree of intricacy is exactly as expected for semi-realistic models that might describe the real world.

- Studies of such models, even though unstable, can eventually shed light on the degree to which vacuum stability affects other phenomenological properties.

- N=0 string models may provide an alternative means of understanding our N=0 world, thus worth understanding in their own right.

- It's fun.

So let's proceed...

The models

- Four-dimensional weakly-coupled heterotic strings
- Realized through the free-fermionic construction:
  - Worldsheet (super)CFT with c=(9,22) realized in terms of free complex NS or Ramond fermions
  - Different spin-structures contribute to partition function with GSO phases preserving modular invariance and guaranteeing proper spin-statistics relations
- Models generated through random but self-consistent choices of fermion boundary conditions (R or NS) and spin-structure phases
- Models restricted to tachyon-free N=0 for now, generalized to N=0,1,2,4 later
- *Complex* fermions only
  - No rank-cutting: all gauge groups rank=22, simply laced

Advantages of the fermionic construction:

Kawai, Lewellen, Tye; Antoniadis, Bachas, Kounnas

- Relatively easy to generate models with an intricacy and complexity that is hard to duplicate through more geometric constructions --- indeed, through sequential layers of twists and projections, can easily generate models for which no geometric interpretation is apparent.
- Substantial overlaps with Narain (bosonic) lattice formulations and orbifold/Wilson-line constructions.
- Although reaches only discrete points in full model space, such points tend to represent the models of most phenomenological relevance (e.g., containing non-abelian gauge groups).
- Full tree-level spectrum and couplings calculable.
- Straightforward to automate for computer searches.

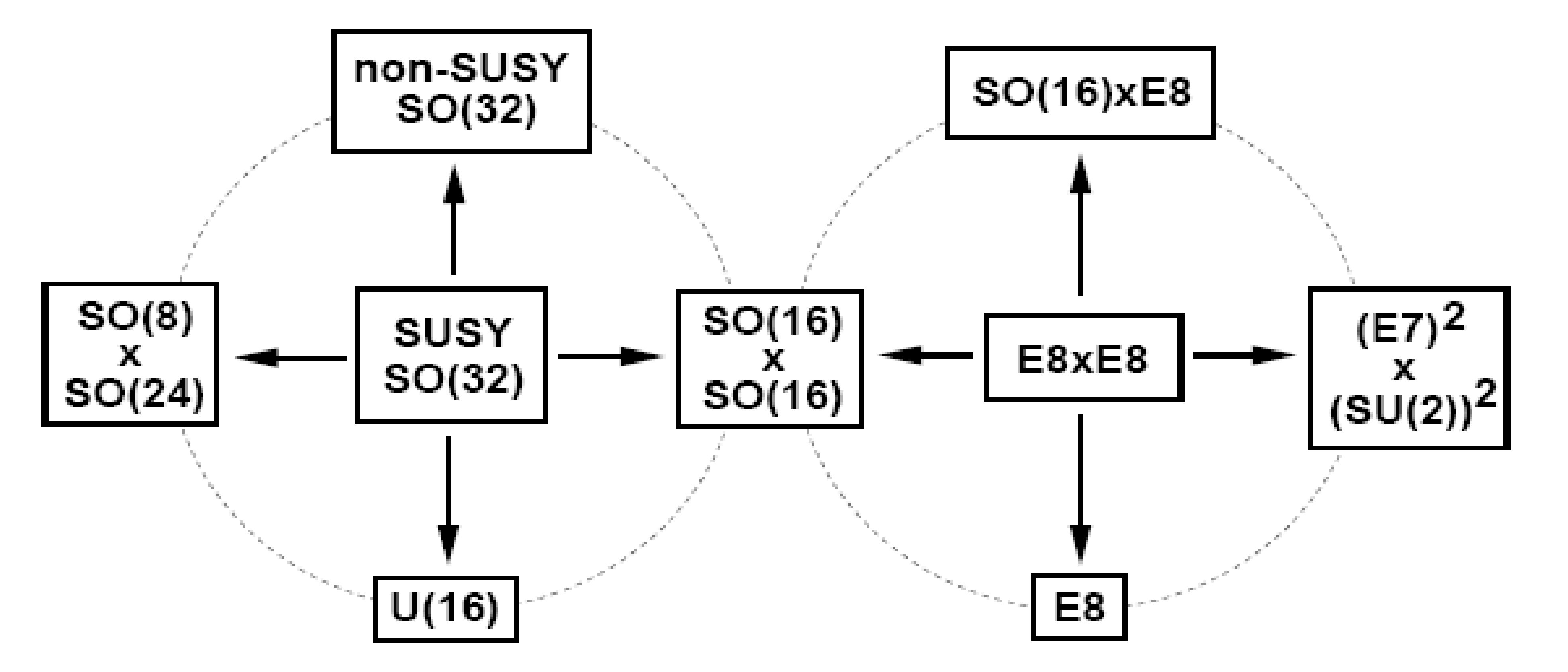
Indeed, models constructed using these techniques span almost the entire spectrum of closed-string models...

- SM and MSSM-like models
- String GUT models
- Models with and without exotic chiral matter, etc.

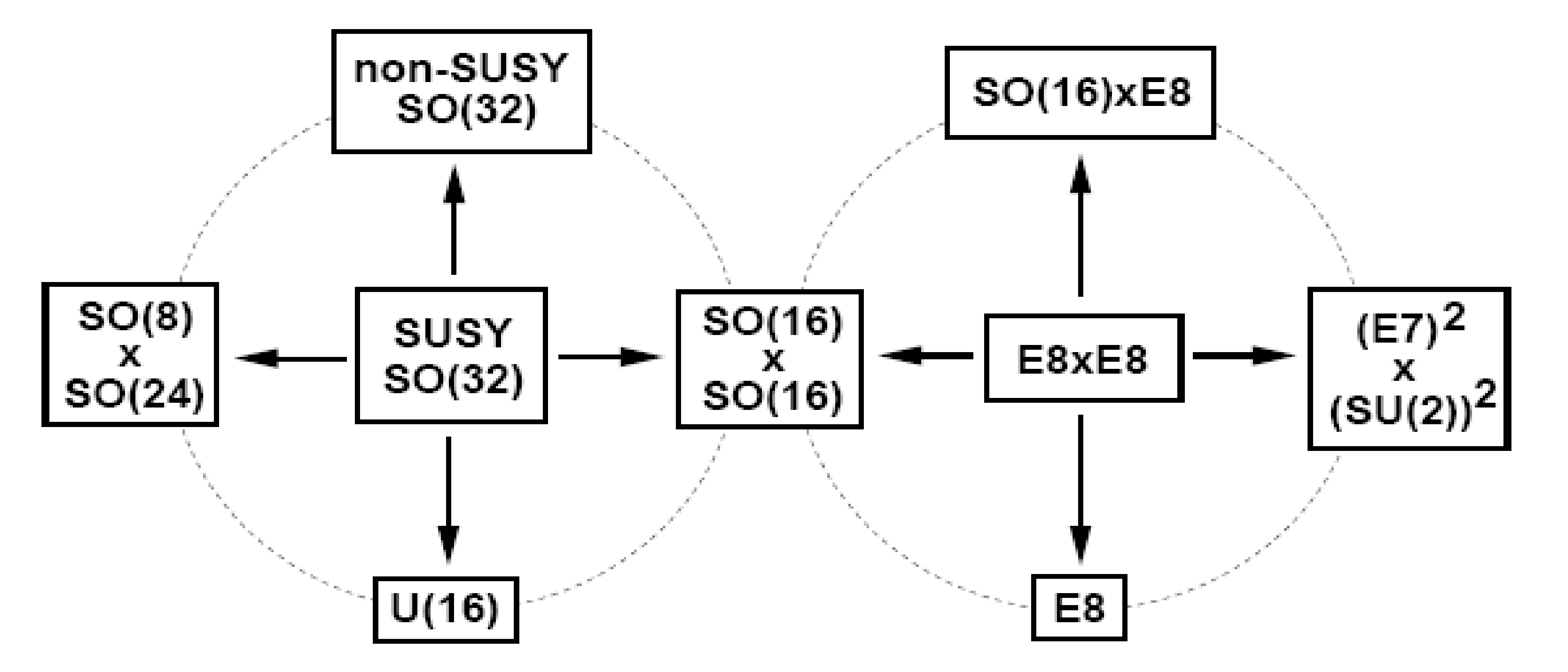
Indeed, despite the abstract mathematical power of more geometric (CY) formulations, most detailed work in closed-string model-building over the past two decades has occurred through free-field (bosonic/fermionic/orbifold) constructions.

First, a warm-up exercise...

The heterotic “landscape” in D=10



- Nine distinct models, connected through orbifold “web”
- Two models with N=1 SUSY, one N=0 but tachyon-free, six N=0 and tachyonic
- Large variation in gauge groups
- Ranks ≤ 16



Lessons from D=10...

- Only some gauge groups are realizable, not all
- Correlations can exist between seemingly unrelated features (e.g., SUSY and gauge group)
- Our model-construction techniques will not reach every model.

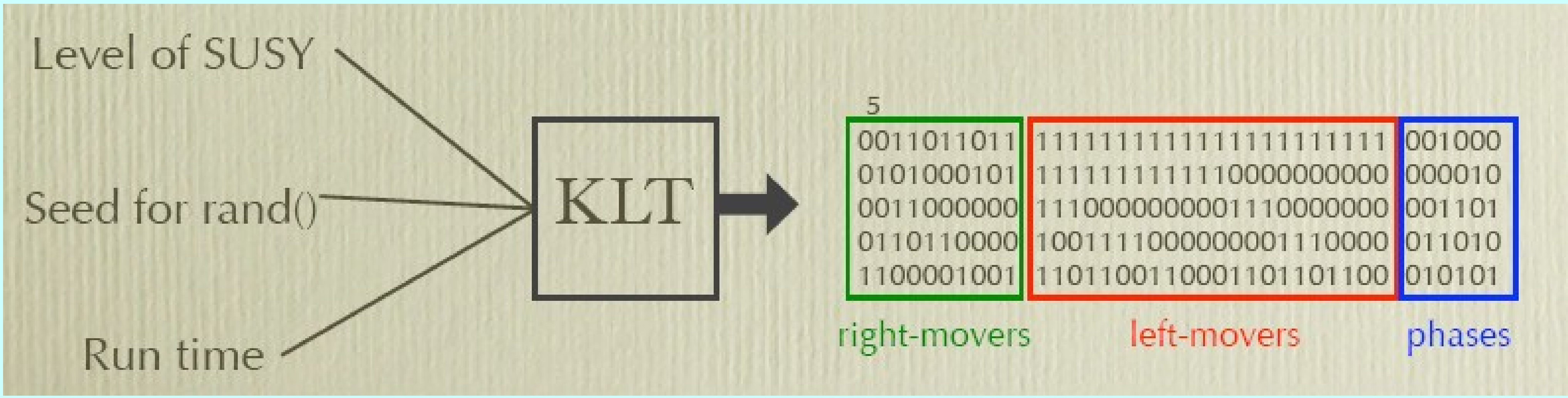
Now turn to D=4.
Landscape is vastly more complex…

- Many, many models in this class alone.
- Orbifold twists/Wilson lines can be extremely complicated, can act sequentially in non-trivial overlapping ways with fairly complicated patterns of simultaneous GSO projections
- Now can have N=1,2,4 SUSY as well as N=0 (both with and without tachyons) --- first focus on N=0 tachyon-free models
- Rank of gauge groups ≤ 22 --- in this study, rank=22 only.

Need to generate models randomly and analyze them systematically.

How we do it:

Step #1: Generating models

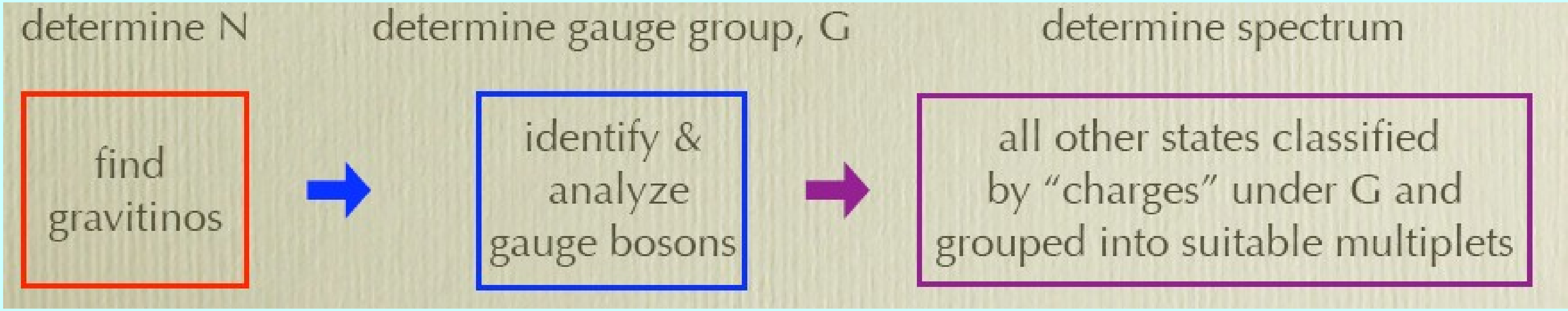


Can generate millions/billions of self-consistent configurations of twists/phases very easily!

D. Senechal

Step #2: Analyze candidate model

- For each spin-structure, enumerate all states in Fock space satisfying level-matching and GSO constraints
- Organize these states into meaningful representations
    - first gravitinos, then appropriate gauge multiplets, finally rest of spectrum



Resulting spectrum is then quoted in terms of Dynkin labels and U(1) charges, labelled as real or complex, chiral or non-chiral, etc.

```
Supersymmetry N = 0

57 gauge bosons in SU(4) x SU(2)^14 x U(1)^5

34 Fermions irreps:
24     010 1 1 0 0 0 0 0 0 0 0 0 0 0 0   0 0 0 0 0 r
24     010 0 0 0 0 1 1 0 0 0 0 0 0 0 0   0 0 0 0 0 r
24     010 0 0 0 0 0 0 0 0 1 1 0 0 0 0   0 0 0 0 0 r
24     010 0 0 0 0 0 0 0 0 0 0 1 1 0 0   0 0 0 0 0 r
16     000 1 1 1 1 0 0 0 0 0 0 0 0 0 0   0 0 0 0 0 r
16     000 1 1 0 0 0 0 0 0 0 0 0 0 1 1   0 0 0 0 0 r
16     000 0 0 1 1 1 1 0 0 0 0 0 0 0 0   0 0 0 0 0 r
16     000 0 0 1 1 0 0 0 0 1 1 0 0 0 0   0 0 0 0 0 r
...
4      000 1 1 0 0 0 0 0 0 0 0 0 0 0 0   0 0 0 0 2 c
4      000 0 0 0 0 1 1 0 0 0 0 0 0 0 0   0 0 0 0 2 c
4      000 0 0 0 0 0 0 1 1 0 0 0 0 0 0   0 2 0 0 0 c
4      000 0 0 0 0 0 0 1 1 0 0 0 0 0 0   0 0 0 2 0 c
...
35 Scalar irreps:
24     010 0 0 1 1 0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
...
```

D. Senechal

Important point: Many different configurations of orbifold twists/Wilson lines give rise to exactly the same particle spectrum in spacetime!
A given string model can have multiple realizations in terms of the underlying construction.

Thus, must analyze the spectrum of each candidate model *and compare with those of all previous models* before deciding whether a new, distinct string model has truly been found.

This is computationally intensive, and turns out to be the limiting factor in studies of this type.

(more about this later...)

So what do we find?

First, discuss statistical results from a sample of $\sim 10^5$ non-supersymmetric, tachyon-free models...

KRD, hep-th/0602286

Since all models have rank 22, it turns out that one important way to categorize models is according to their numbers of irreducible gauge group factors.

How “shattered” (or “twisted”) is the total gauge group?

Models range from

$G = SO(44)$ — Least “shatter”

[analogue of SO(32) in D=10]

↕

$G = U(1)^n \times SU(2)^{22-n}$ — Most “shatter”

Models fill out a "tree" when arranged as a function of shatter…

f=1: SO(44) only – unique model is "root" of tree

f=2: 34 distinct models, 4 unique gauge groups of form G= SO(44-n) x SO(n) for n=8,12,16,20

f=3: 186 distinct models, 8 distinct gauge groups. This is the first shatter level at which E groups appear, always E8.

f=4: Thousands of distinct models, 34 distinct gauge groups. First appearance of E7, U(1), and SU(n) with n=4,8,12

Models fill out a "tree" when arranged as a function of shatter...

f=5: 34 distinct gauge groups. First appearance of E6 and SU(n) with n=6,10,14.

f=6: 70 distinct gauge groups. First appearance of SU(7). *Disappearance* of SU(14).

f=7: 75 distinct gauge groups, only one with E8. First appearance of SU(5).

f=8: 89 distinct gauge groups. Disappearance of E8. First appearance of SU(3).

-- thus, e.g., no models with E8 x SU(3) x ... !

Models fill out a "tree" when arranged as a function of shatter...

f=22: $G = U(1)^n \times SU(2)^{22-n}$ only

f=21: $G = U(1)^n \times SU(2)^{20-n} \times SU(3)$ only

--- SU(3) factor *must* appear

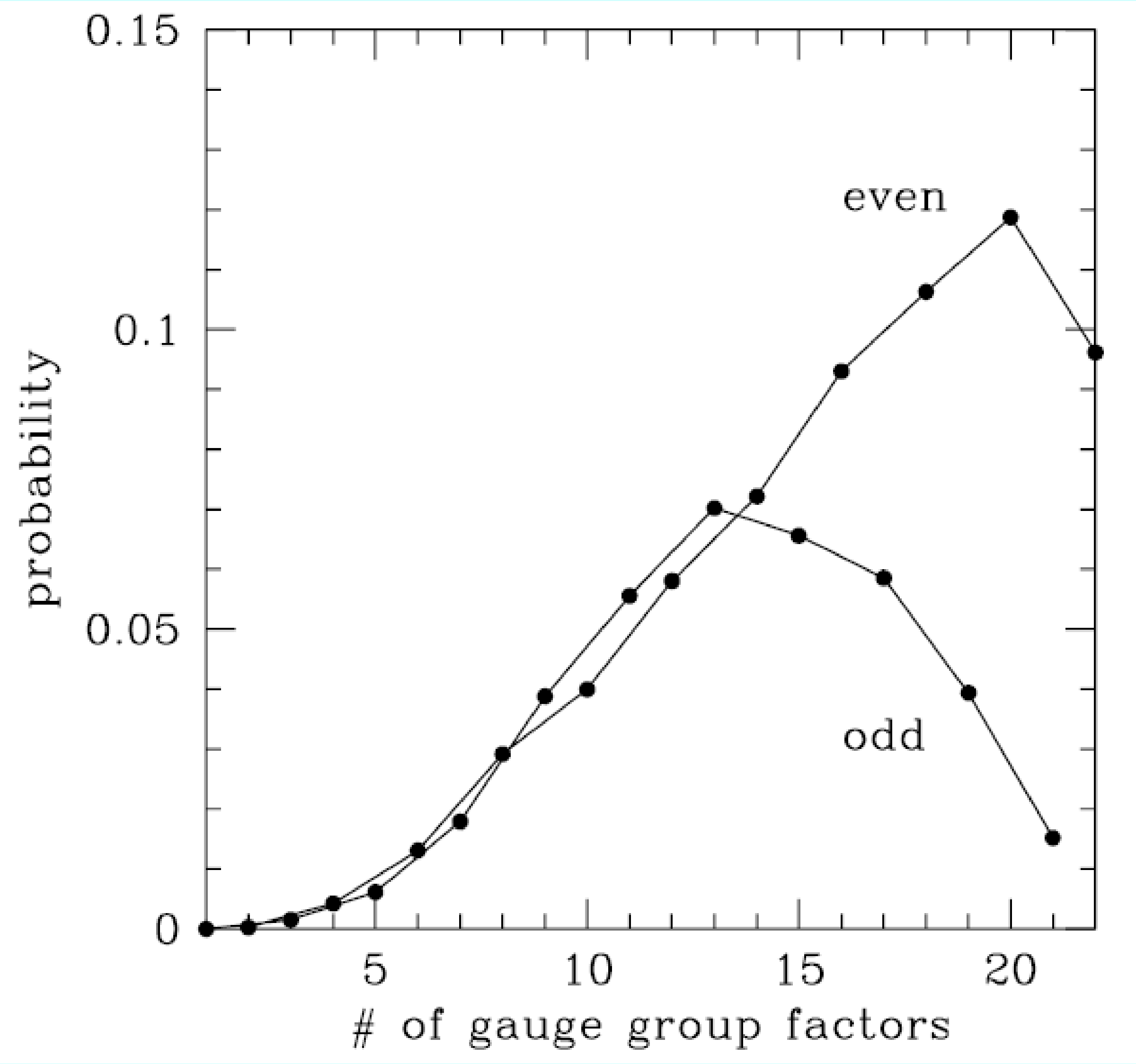
f=20: 24 distinct gauge groups, all containing either $SU(3)^2$ or $SU(4) \sim SO(6)$

f=19: 37 distinct gauge groups, all containing either $SU(3)^3$ or $SU(3) \times SU(4)$ or $SU(5)$ or $SO(8)$.

In aggregate,

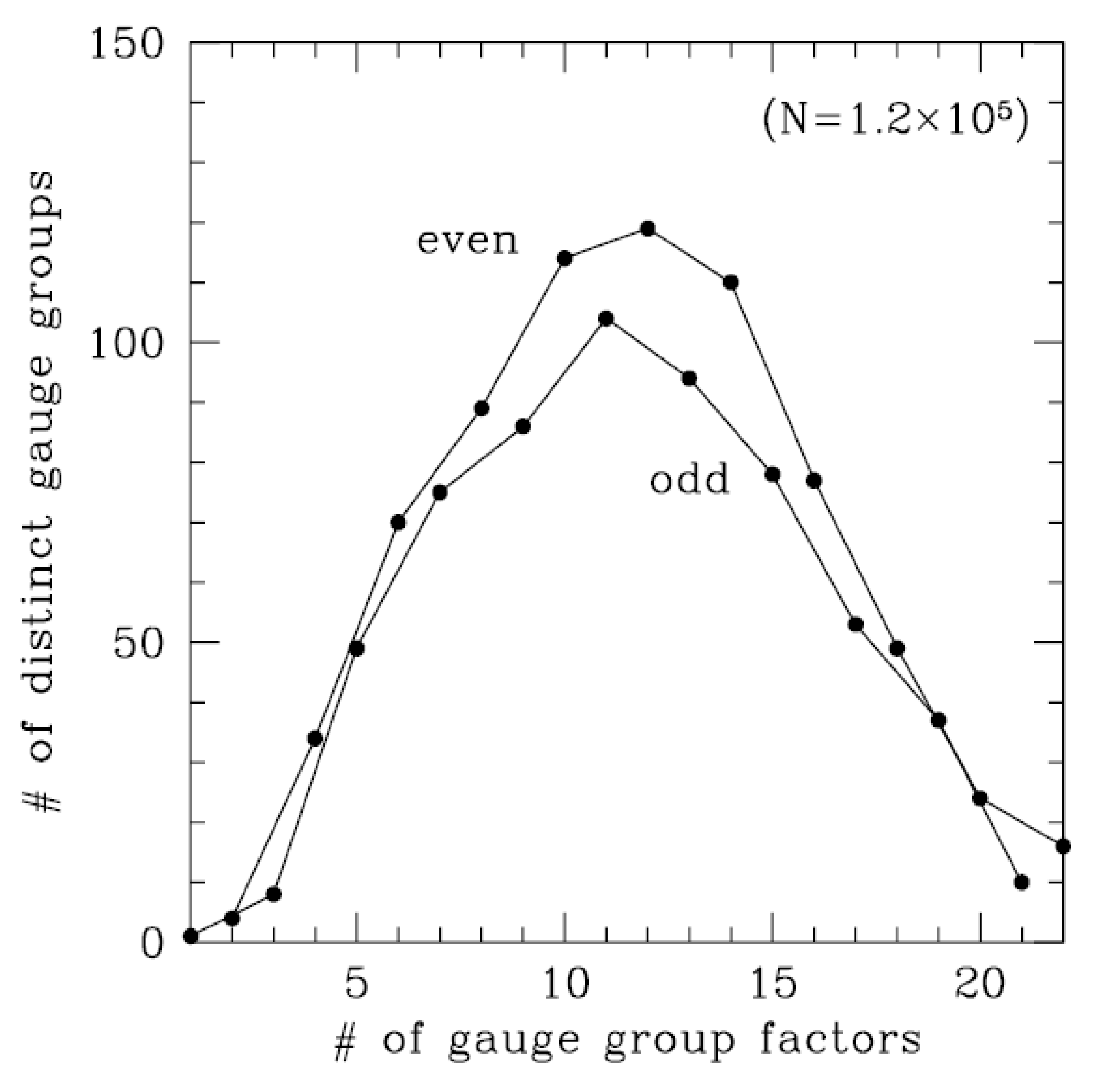
- 123,573 distinct models
- only 1301 distinct gauge groups distributed across all levels of shatter.

Which levels of shatter are most likely?

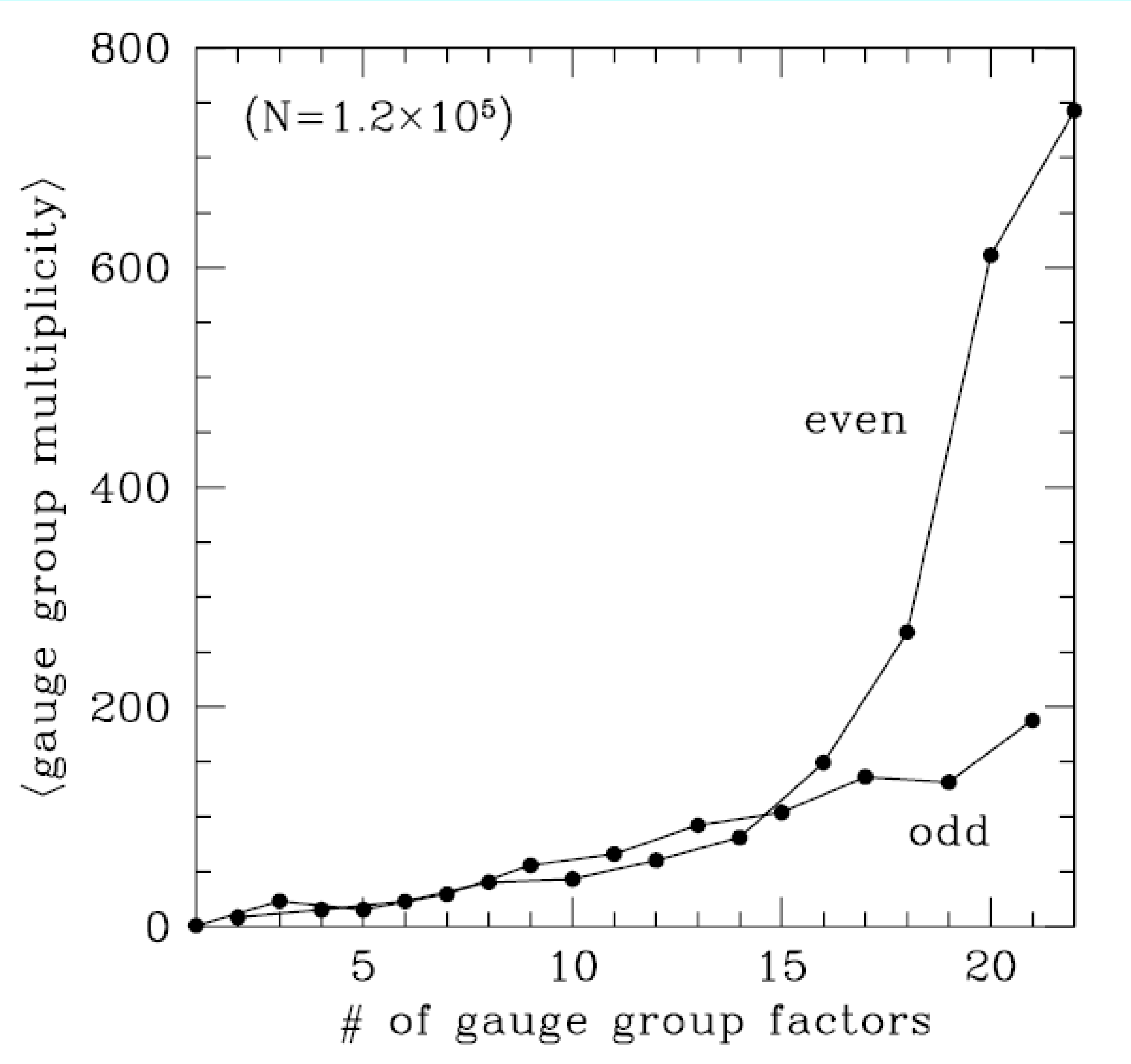


Even shatters appear to dominate at large shatter.

Which levels of shatter give rise to the most distinct *gauge groups* (rather than models)?



Which levels of shatter give rise to the greatest number of distinct string models per gauge group?



Another important issue concerns the *composition* of the gauge groups.

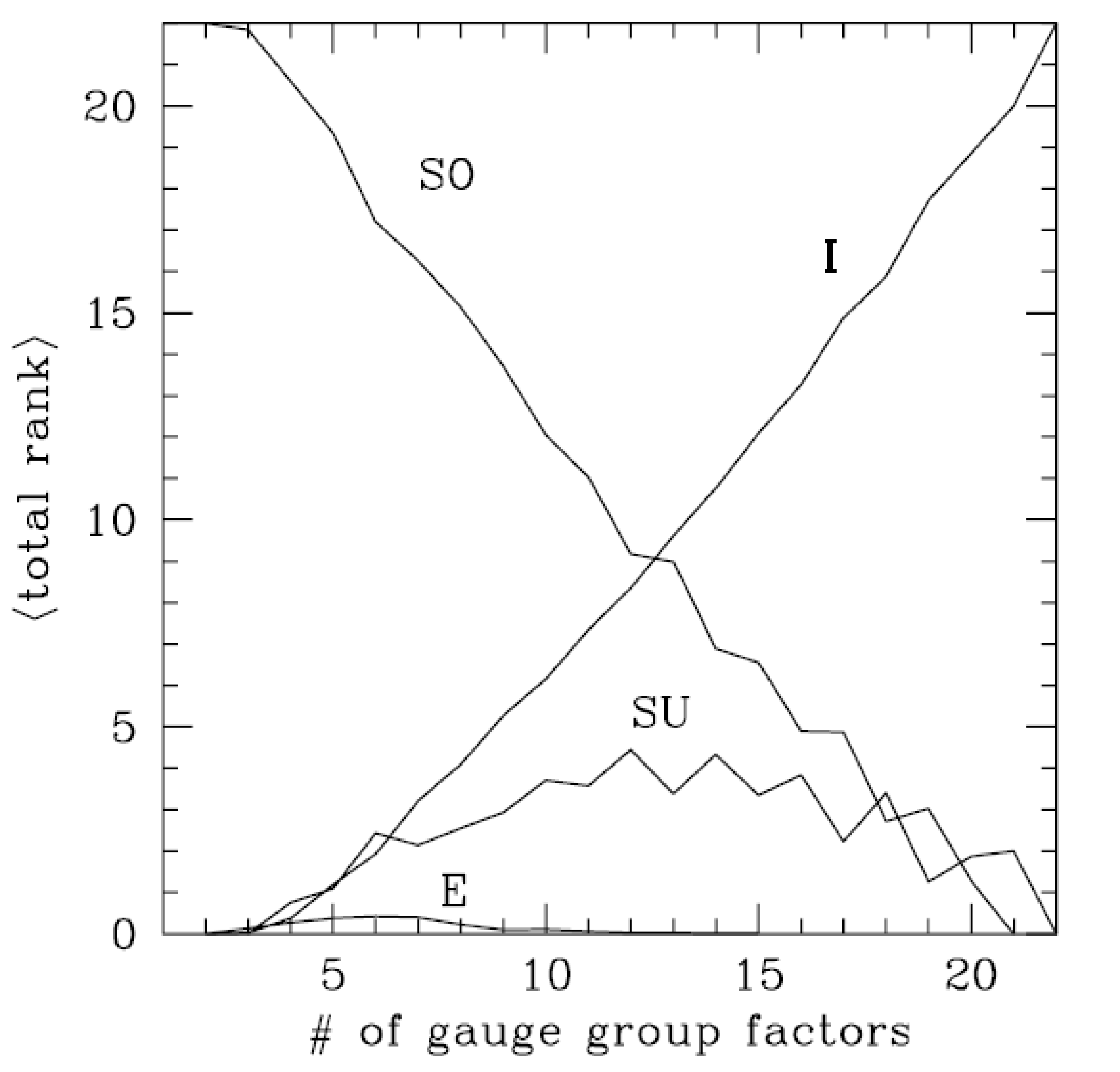
As always, total rank = 22.

How is this total rank distributed amongst

- 'SO' groups
- 'SU' groups
- 'E' groups
- 'I' groups [groups with rank 1: U(1), SU(2)] ?

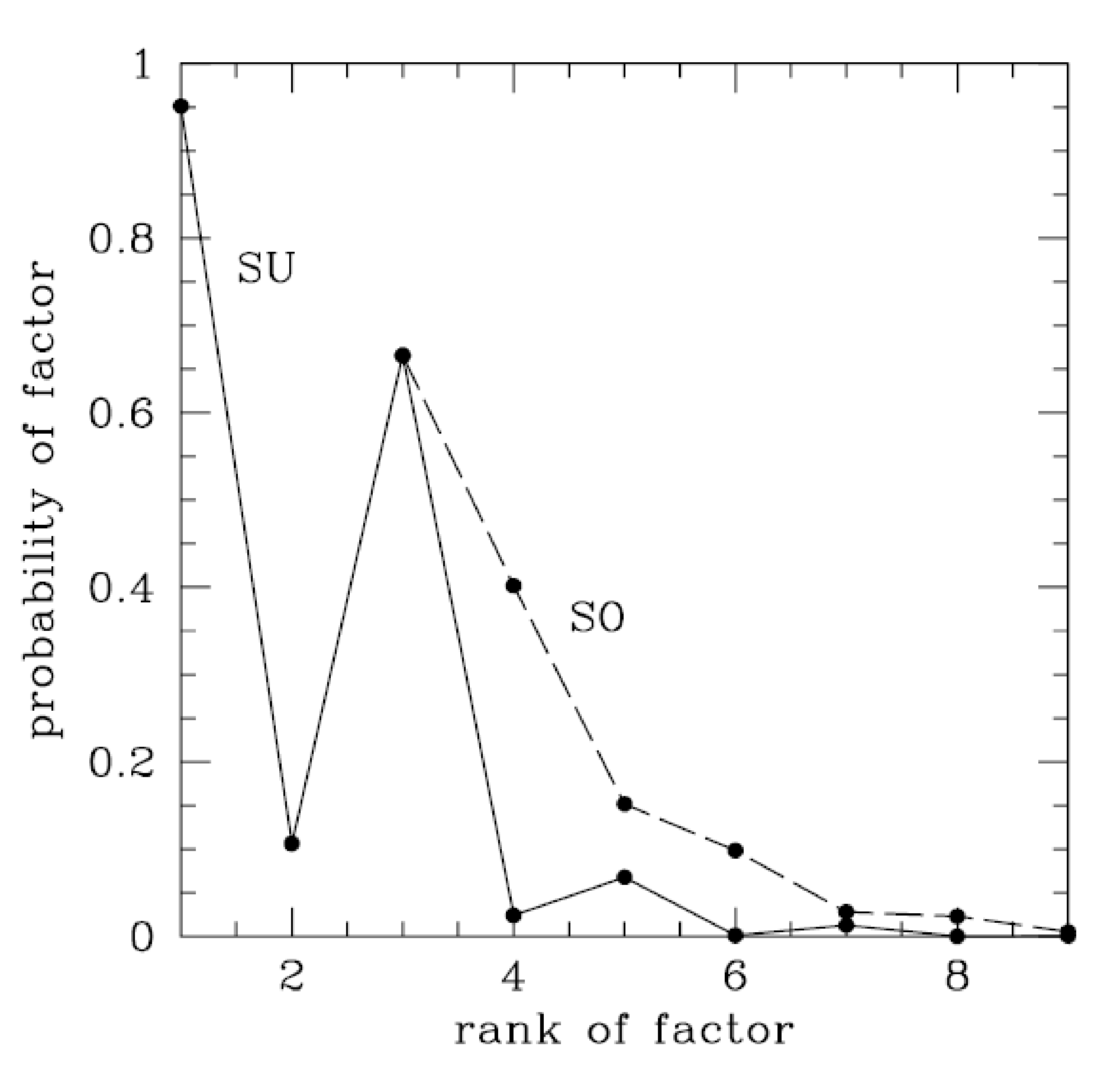
For counting purposes, SU(4)~SO(6) is distributed equally between SO and SU.

As a function of shatter, these individual contributions to the total rank are...



- These are averages over all string models with a given shatter.
- Total of all lines=22.
- Mostly 'SO' for small shatters, mostly 'I' for large shatters.
- 'SU' sizable for intermediate shatters, always =2 for f=21.
- 'E' groups extremely unlikely, given their smallest rank is 6.

How likely are *individual* 'SO' and 'SU' gauge factors?



- These are averages over all string models in the sample.
- SU(2) is ubiquitous, but SU(3) much rarer.
- SU(4)~SO(6), thus two curves coincide at rank=3.
- For larger ranks, 'SO' groups slightly more common than 'SU' groups in our sample.

Indeed, across all string models in our sample,

- 10.65% contain SU(3) factors. Among these models, the average number of such factors is 1.88.
- 95.06% contain SU(2) factors; average number 6.85.
- 90.80% contain U(1) factors; average number 4.40.

By contrast, across all distinct *gauge groups*,

- 23.98% contain SU(3) factors; average number 2.05.
- 73.87% contain SU(2) factors; average number 5.66.
- 91.47% contain U(1) factors; average number 5.10.

Thus, e.g., although SU(3) factors appear in 24% of gauge groups, those groups emerge from actual string models in our sample only half as frequently as we would have expected.

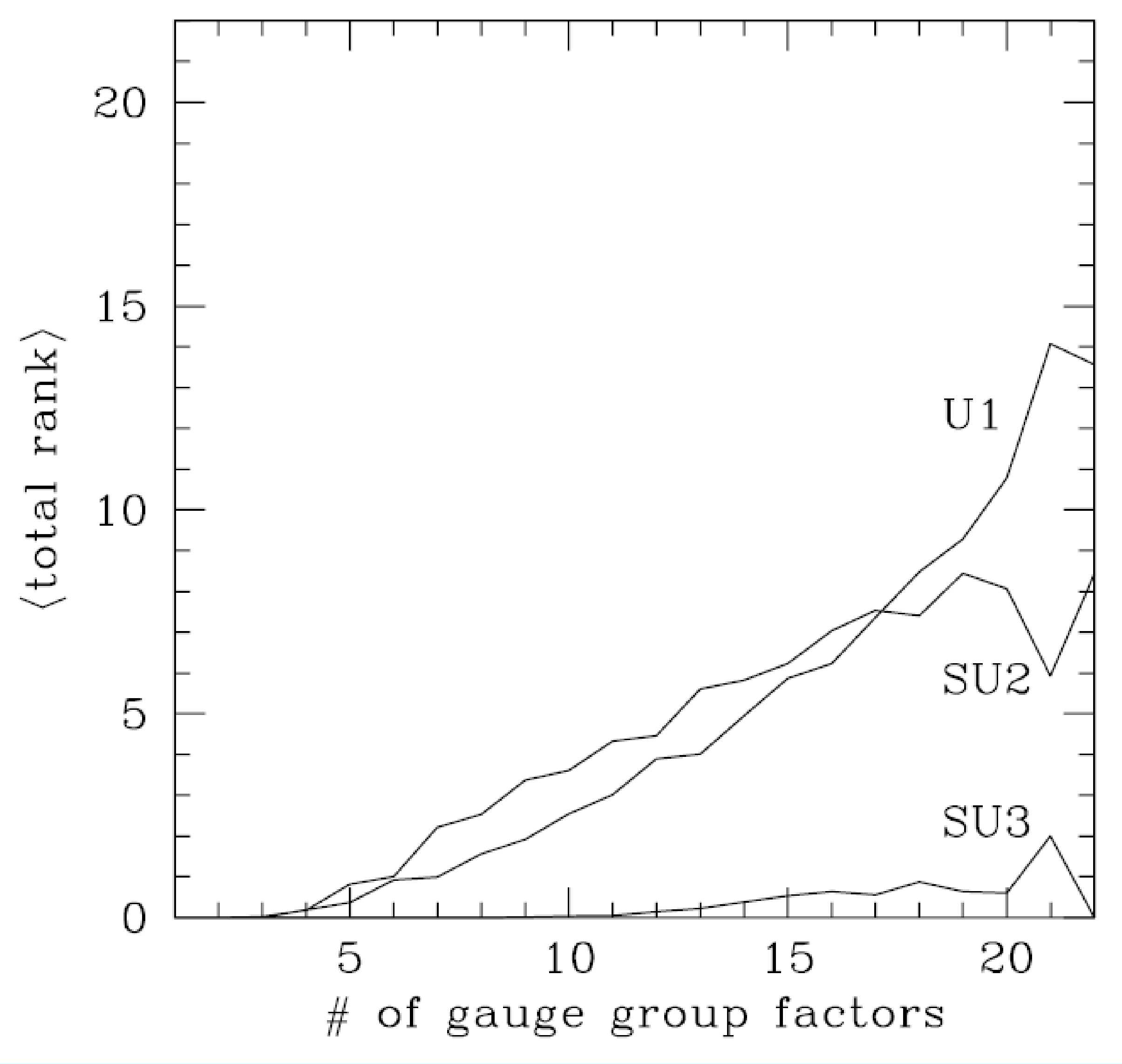
In fact, 99.81% of all heterotic string models in our sample which contain *one or more* SU(n) factors also exhibit an *equal or greater number* of U(1) factors.

True for SU(3) and all SU(n), n≥5.

By contrast, this is true of only 75% of models with SO(2n≥6) factors and only 61% of models with 'E' factors... i.e., no such correlation for these groups!

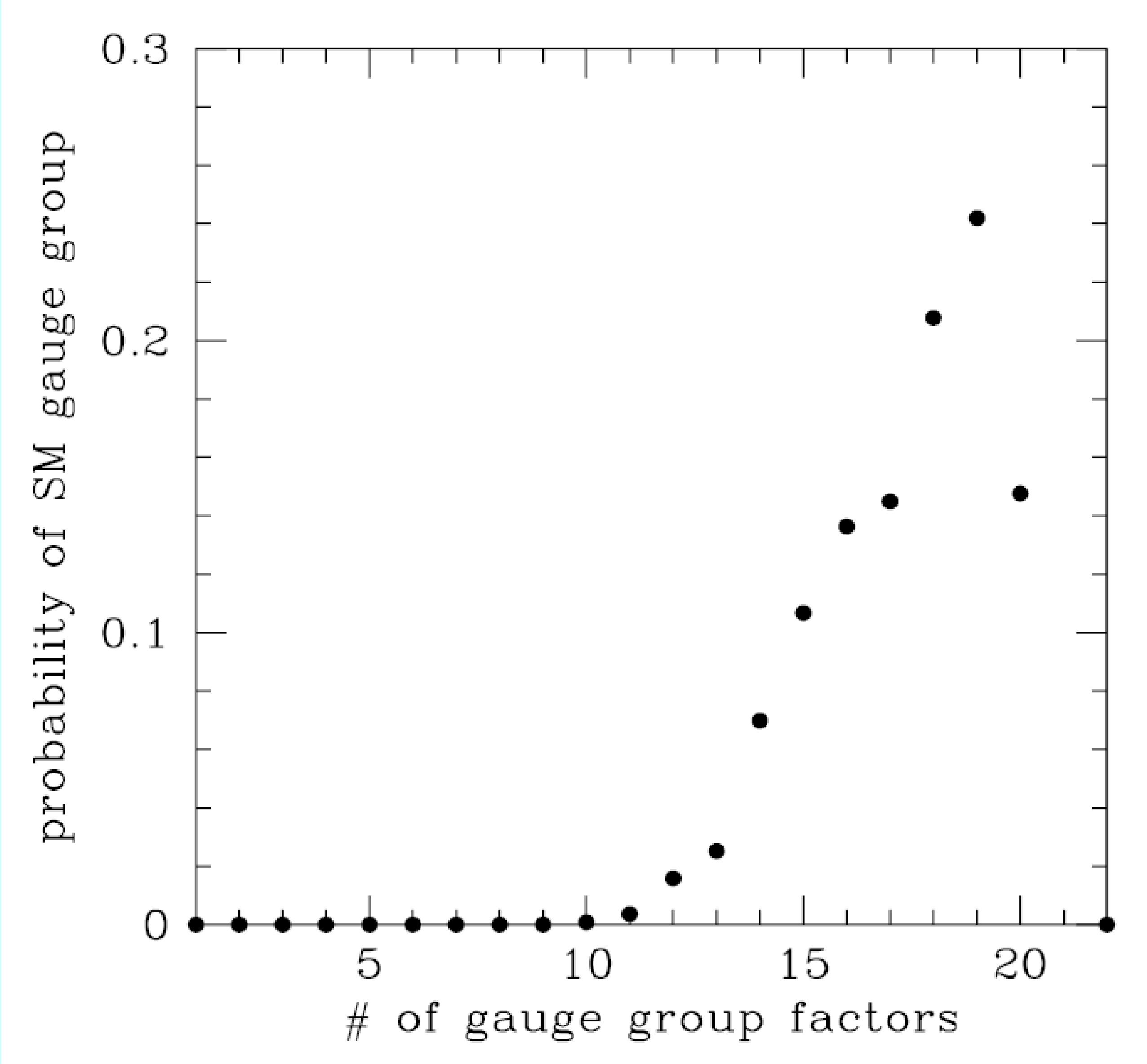
The origin of this SU(n)/U(1) correlation involves the possible embeddings of the charge/momentum lattice on integer/half-integers lattice sites.

How much do SU(3), SU(2), and U(1) individually contribute to the total rank?



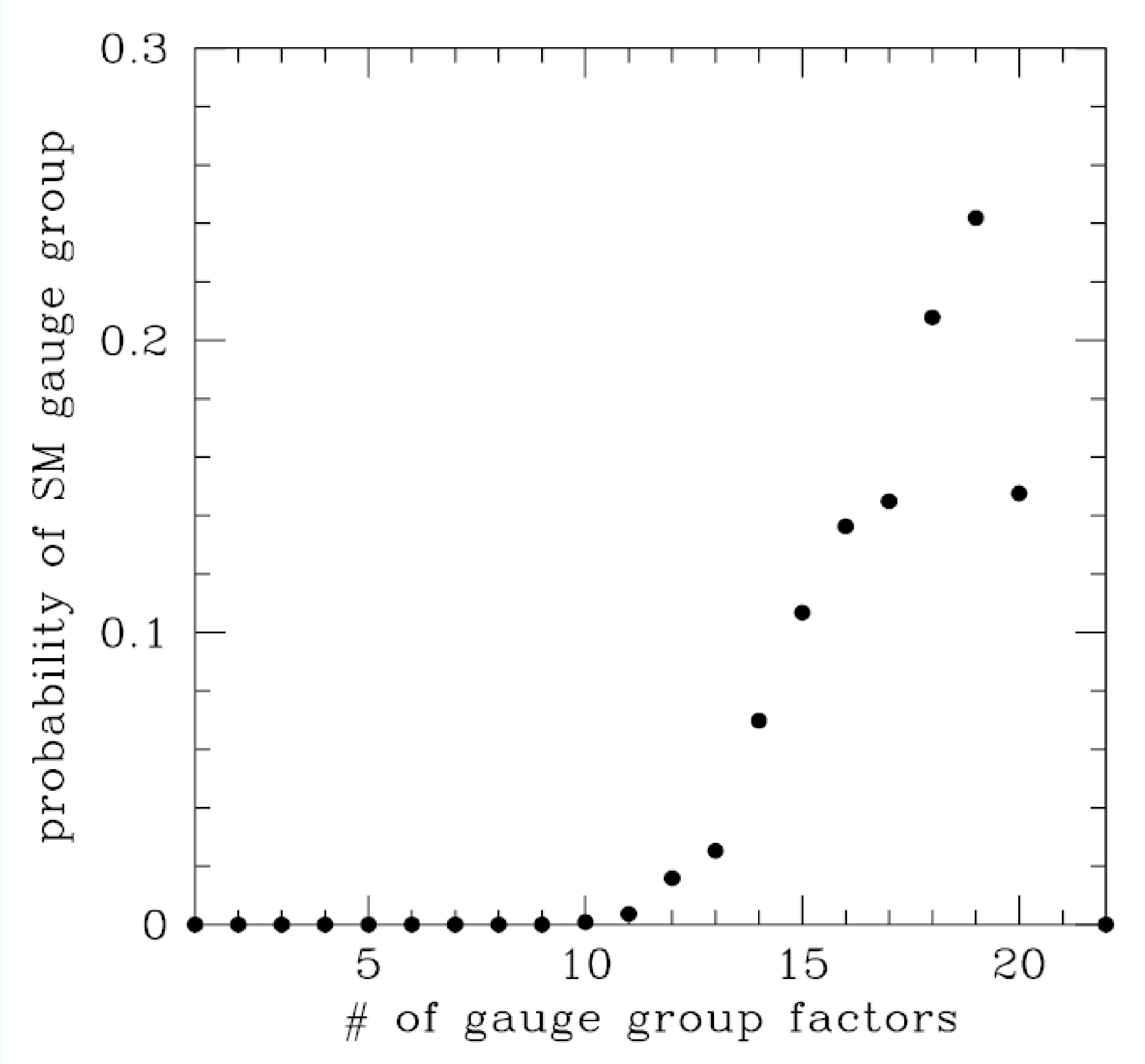
- These factors tend to make sizable contributions only when the shatter is significant.

How likely are SU(3), SU(2), and U(1) to appear *simultaneously* in a given string model in our sample?



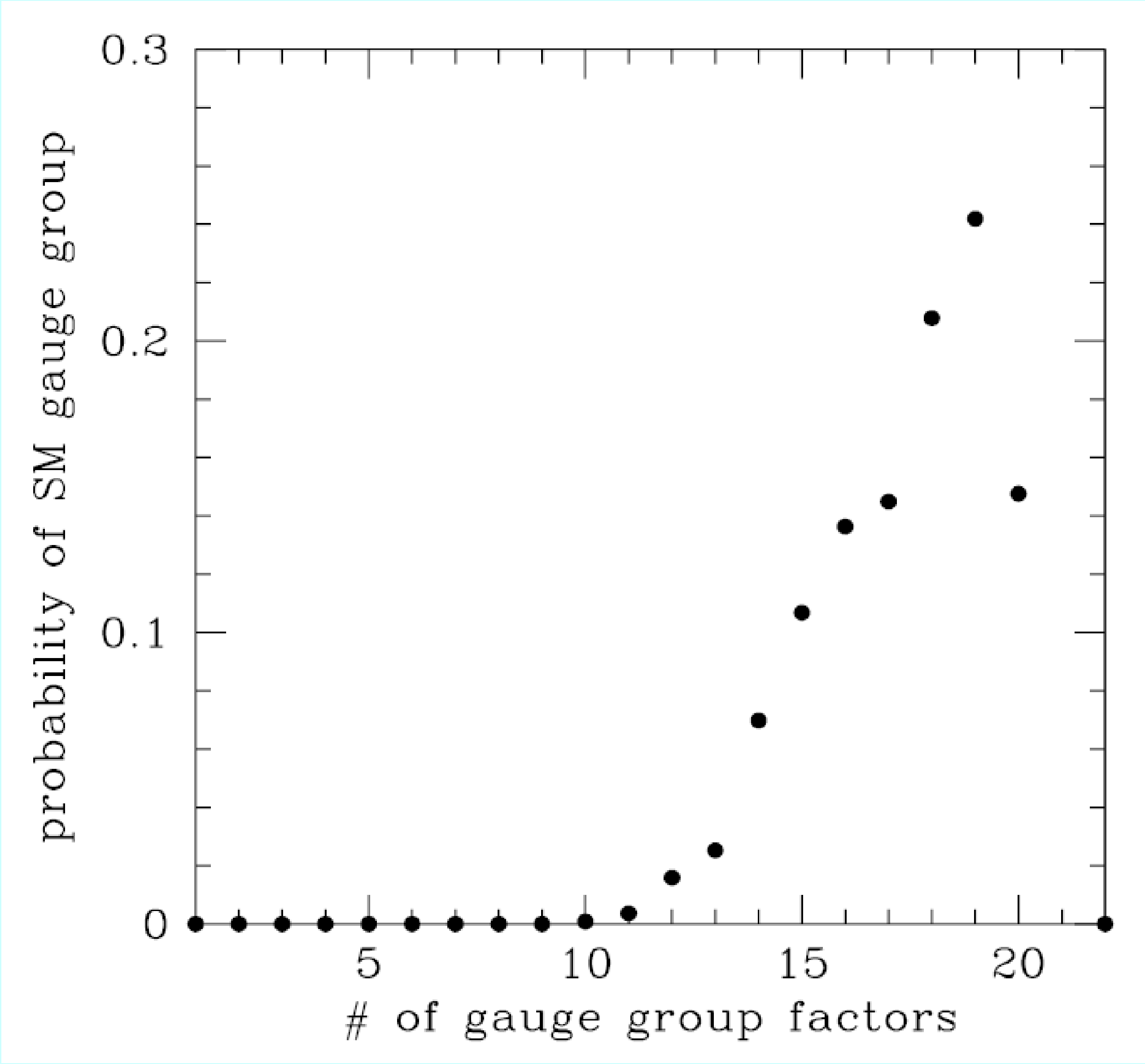
- SM gauge group does not even appear until approximately f>10.
- Probability for SM actually hits
  - 1 for f=21
  - 0 for f=22.
- SM gauge group is most likely to emerge in approximate range between 15 and 21.

How likely are SU(3), SU(2), and U(1) to appear *simultaneously* in a given string model in our sample?



- These conclusions agree with all known such semi-realistic string models in literature.
- Provides limits on possible hidden-sector gauge groups for such models.
- Useful guide for future string model-building.

How likely are SU(3), SU(2), and U(1) to appear *simultaneously* in a given string model in our sample?



Indeed, averaged across all degrees of shatter, the total probability of obtaining the SM gauge group in this sample of models is only 10.05% --- similar to what is found for Type I strings.

How about cross-correlations between *all* possible gauge groups of interest?

What are the joint probabilities that two different gauge group factors will appear within the same string model simultaneously?

This is especially useful to know if one factor is "observable", the other "hidden"...

Correlation probability table (quoted in % of models)...

	U_1	SU_2	SU_3	SU_4	SU_5	$SU_{>5}$	SO_8	SO_{10}	$SO_{>10}$	$E_{6,7,8}$	SM	PS
U_1	87.13	86.56	10.64	65.83	2.41	8.20	32.17	14.72	8.90	0.35	10.05	61.48
SU_2		94.05	10.05	62.80	2.14	7.75	37.29	13.33	12.80	0.47	9.81	54.31
SU_3			7.75	5.61	0.89	0.28	1.44	0.35	0.06	10^{-5}	7.19	5.04
SU_4				35.94	1.43	5.82	24.41	11.15	6.53	0.22	5.18	33.29
SU_5					0.28	0.09	0.46	0.14	0.02	0	0.73	1.21
$SU_{>5}$						0.59	3.30	1.65	1.03	0.06	0.25	4.87
SO_8							12.68	6.43	8.66	0.30	1.19	22.02
SO_{10}								2.04	2.57	0.13	0.25	9.44
$SO_{>10}$									3.03	0.25	0.03	5.25
$E_{6,7,8}$										0.01	0	0.13
SM											7.12	3.86
PS												26.86
total:	90.80	95.06	10.64	66.53	2.41	8.20	40.17	15.17	14.94	0.57	10.05	62.05

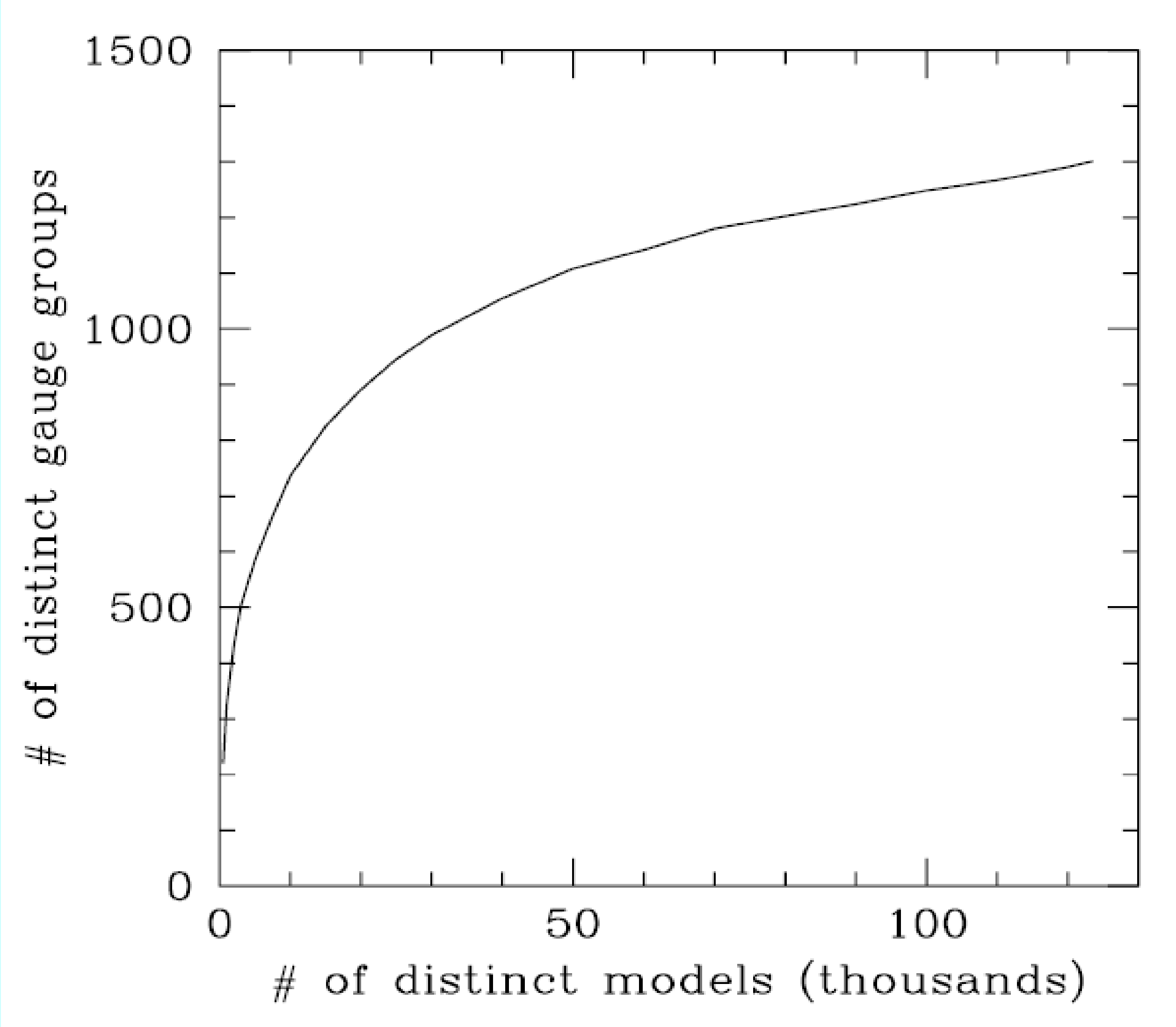
- SM = Standard Model;   PS = Pati Salam SO(4) x SO(6)
- Off-diagonal entries show pairwise percentages;
  diagonal entries show percentages for factor appearing *twice*.
- "Total" is *uncorrelated* probability for single group factor.

Correlation probability table (quoted in % of models)...

	U_1	SU_2	SU_3	SU_4	SU_5	$SU_{>5}$	SO_8	SO_{10}	$SO_{>10}$	$E_{6,7,8}$	SM	PS
U_1	87.13	86.56	10.64	65.83	2.41	8.20	32.17	14.72	8.90	0.35	10.05	61.48
SU_2		94.05	10.05	62.80	2.14	7.75	37.29	13.33	12.80	0.47	9.81	54.31
SU_3			7.75	5.61	0.89	0.28	1.44	0.35	0.06	10^{-5}	7.19	5.04
SU_4				35.94	1.43	5.82	24.41	11.15	6.53	0.22	5.18	33.29
SU_5					0.28	0.09	0.46	0.14	0.02	0	0.73	1.21
$SU_{>5}$						0.59	3.30	1.65	1.03	0.06	0.25	4.87
SO_8							12.68	6.43	8.66	0.30	1.19	22.02
SO_{10}								2.04	2.57	0.13	0.25	9.44
$SO_{>10}$									3.03	0.25	0.03	5.25
$E_{6,7,8}$										0.01	0	0.13
SM											7.12	3.86
PS												26.86
total:	90.80	95.06	10.64	66.53	2.41	8.20	40.17	15.17	14.94	0.57	10.05	62.05

- Almost all SU(3), SU(n≥5) factors come with U(1), as already noted.
- No models with SU(5) x (any E-group); no models with SM x (E-group); only one with SU(3) x (E-group).
- Overall, Pati-Salam is *much more prevalent* than SM, while SO(10) is *somewhat more prevalent* and SU(5) is *slightly less prevalent* than SM.

Finally, we have seen that we found only ~1300 distinct gauge groups for over ~120,000 models. *How does the number of gauge groups grow with the number of models constructed?*



- Gets harder and harder to find new gauge groups as we continue to generate models.
- Curve appears to *saturate* at a maximum number of possible heterotic gauge groups…?

Now we turn to the one-loop vacuum energy densities (cosmological constants) λ associated with these models.

Define:

$$\Lambda \equiv \int_{\mathcal{F}} \frac{d^2\tau}{(\mathrm{Im}\,\tau)^2} Z(\tau)$$

where $Z(\tau)$ is the one-loop partition function:

$$Z(\tau) \equiv \mathrm{Tr}\,(-1)^F\,\overline{q}^{H_R}\,q^{H_L}$$

Then

$$\lambda \equiv -\tfrac{1}{2}\mathcal{M}^4\Lambda$$

where M is the reduced Planck scale.

Since these models are non-supersymmetric, we generally expect non-zero Λ.

Of course, just as with the ten-dimensional SO(16) x SO(16) heterotic string, the presence of a non-zero Λ indicates that these models are unstable beyond tree level.

Why then focus on this quantity?

Warning: brief editorial ahead.

- Simplest possible one-loop amplitude one can calculate --- its properties may hold lessons for more complicated amplitudes. For example, general n-point amplitudes are related to this amplitude through differentiations (e.g., string threshold corrections). Similar expressions thus apply even for SUSY theories.
- By examining the behavior of such stringy amplitudes, gain insight into the extent to which effective supergravity calculations might hold in the full string context. (E.g., how great a contribution do massive string states actually make?)
- When correlated with information about gauge groups, may learn how gauge groups influence the sizes of these sorts of amplitudes.
- Finally --- and most importantly --- values of Λ relate directly back to fundamental questions of SUSY-breaking and vacuum stability. If we can find models with $\Lambda = 0$ (even approximately), we will have found good approximations to stable vacua with broken SUSY.

We now rejoin the mathematics, already in progress...

First, note that when we evaluate the partition function for a given string model, we obtain a double-power series:

$$Z(\tau) \;=\; \tau_2^{-1} \sum_{mn} b_{mn}\, \overline{q}^m\, q^n$$

Net numbers of bosons minus fermions with left/right worldsheet energies (*m*,*n*).

Change variables to

- sum $s = m+n$ (total WS energy ~ mass^2)
- difference $d = |m-n|$ ("off-shell" amount)

and accordingly, define new degeneracy matrix:

$$a_{sd} \;\equiv\; b_{(s-d)/2,(s+d)/2} \;+\; b_{(s+d)/2,(s-d)/2}$$

We can then write the one-loop vacuum amplitude as

$$\Lambda = \sum_{s,d} a_{sd}\, I_{sd}$$

where

$$I_{sd} \equiv \int_{\mathcal{F}} \frac{d^2\tau}{\tau_2^3} \exp(-2\pi s\tau_2) \cos(2\pi d\tau_1)$$

contribution to Λ from each state with total WS energy s and "off-shell" degree d.

Thus, for each heterotic string model, we simply calculate the net degeneracies of bosons minus fermions in the spectrum for each relevant (s,d) and then tally the above sum to find Λ.

But note the magnitudes of these individual contributions:

s	d	I_{sd}	s	d	I_{sd}
−1.0	1	− 12.192319	1.0	0	0.000330
−0.5	1	−0.617138	1.0	1	−0.000085
0.0	0	0.549306	1.0	2	0.000035
0.0	1	−0.031524	1.0	3	−0.000018
0.0	2	0.009896	1.5	0	0.000013
0.5	0	0.009997	1.5	1	−0.000004
0.5	1	−0.001626	1.5	2	0.000002
0.5	2	0.000587	1.5	3	−0.000001

The biggest contributions are from *off-shell tachyons* with

$(s,d) = (-1,\ 1)$ ==> $(m,n) = (0,\ -1)$!

These are *22 times larger* than from massless on-shell states!

Do such $(m,n)=(0,-1)$ states actually exist in the non-SUSY string spectrum?

Yes!

As long as there's a graviton, there's also:

proto-graviton: $\tilde{b}^{\mu}_{-1/2}|0\rangle_R \otimes |0\rangle_L$

... same as the graviton but without coordinate excitation.

Always there for same reason that graviton is always there!

Try to cancel its contributions with a similar fermion?

proto-gravitino: $\{\tilde{b}_0\}^{\alpha}|0\rangle_R \otimes |0\rangle_L$.

... same as the *gravitino* but without coordinate excitation.

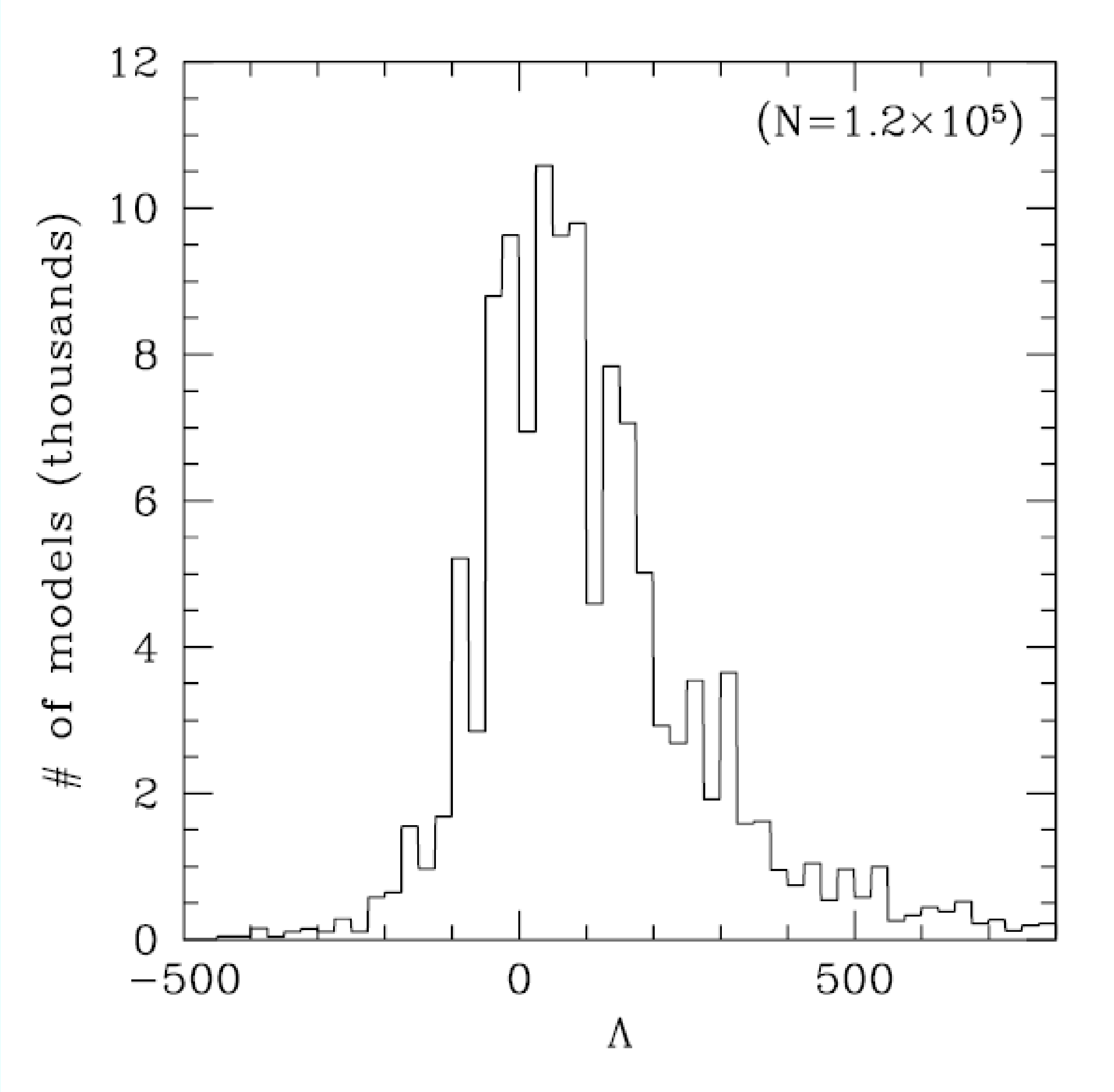
Always projected out when gravitinos are projected out!

Upshot:

- Even before massless states are considered, uncancelled off-shell (bosonic) tachyonic string states produce a significant bias towards $\Lambda < 0$ (i.e., $\lambda > 0$).
- Although contributions from massive states fall exponentially, their numbers *grow* exponentially. *Contributions all the way up to 5^{th} or 6^{th} mass levels are also significant.*

Thus, contributions from the infinite towers of string states, both on-shell and off-shell, are critical for determining not only the magnitude but also the sign of the one-loop cosmological constant. Examination of the massless string spectrum (e.g., through effective low-energy field-theory analysis) is not enough.

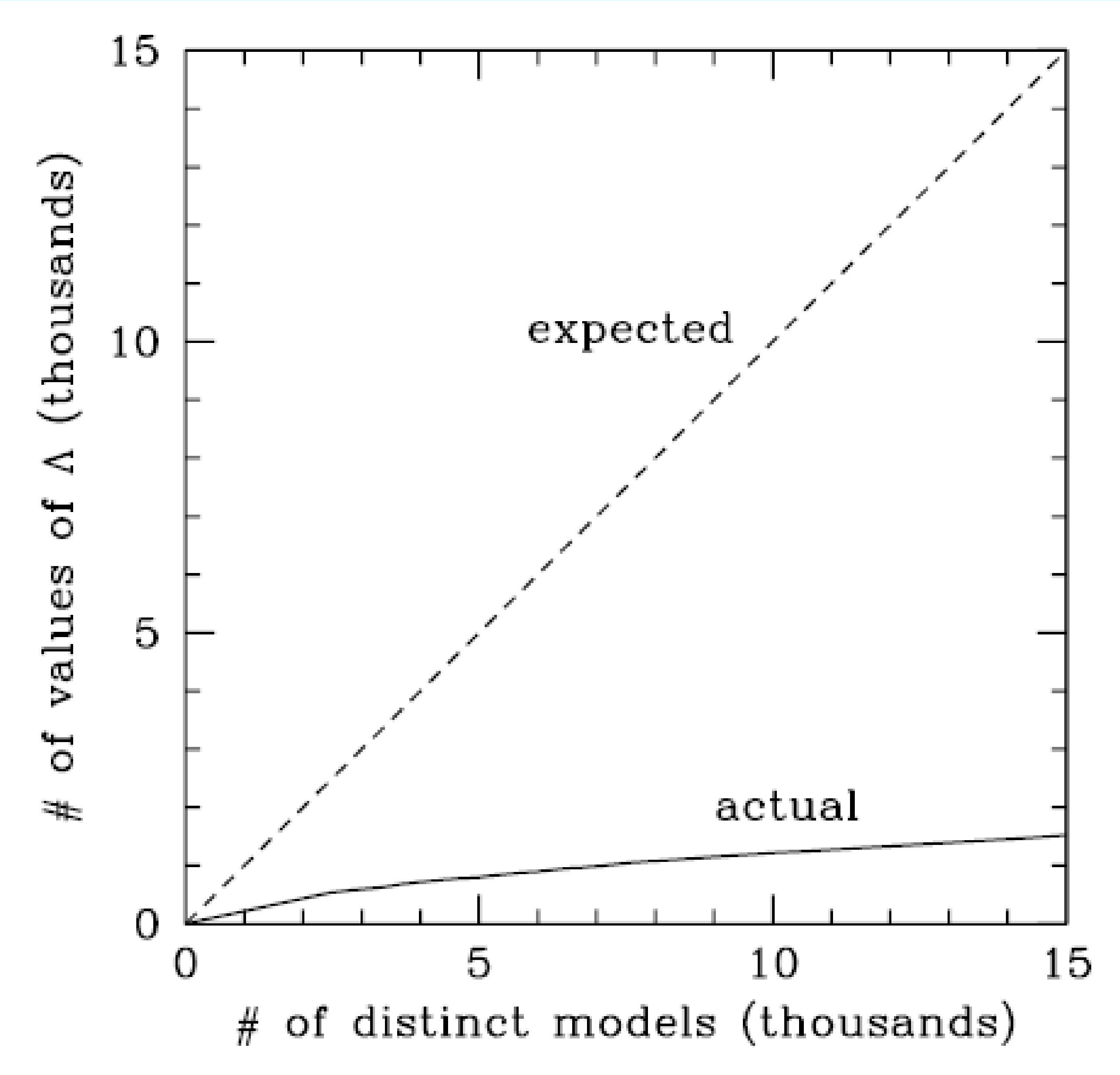
So what values of Λ do we find for our sample?



- Both positive and negative values emerge, with over 73% *positive* (i.e., *negative* λ --> AdS).
- Over 10^5 models, but smallest value of $|\Lambda|$ found is 0.0187.

Why none smaller?

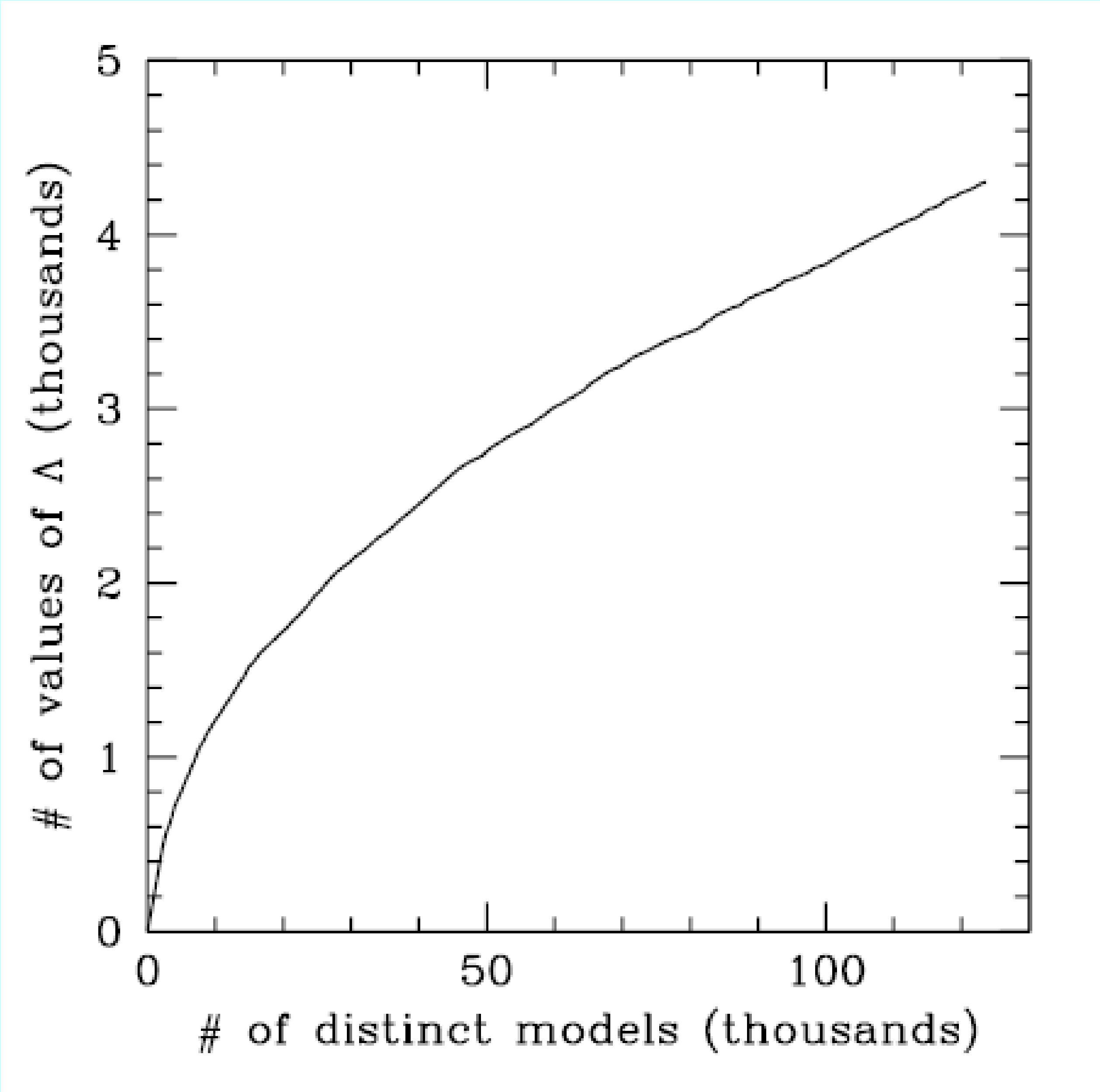
There's a great redundancy in values of Λ !



- The number of values of Λ found is significantly less than the number of models examined!
- Unrelated models with completely different gauge groups and particle content can nevertheless have *identical* values of Λ !

There's a great redundancy in values of Λ !

This trend becomes increasingly severe as we generate more and more models…



Why does this happen?

85% of the degeneracies

- The number of possible modular-invariant partition functions is significantly smaller than the number of distinct string models which can be constructed.
- Two different partition functions can differ by a function which is purely imaginary (makes no contribution to Λ).
- Two different partition functions can differ by an expression proportional to the Jacobi factor

$$J \equiv \frac{1}{\eta^4}\left(\vartheta_3^4 - \vartheta_4^4 - \vartheta_2^4\right) = 0$$

15% of the degeneracies

- Two different partition functions can differ by an expression which has an Atkin-Lehner symmetry (i.e., non-zero, but *integrates* to zero over the fundamental domain).

(Moore, 1987; KRD, 1990)

In fact, it appears that the number of cosmological constant values may actually *saturate*...

If so, fit curve to exponential form

$$\Sigma(t) = N_0 \left(1 - e^{-t/t_0}\right)$$

maximum value (for N_0)

"time constant" (for t_0)

find $N_0 \sim 5500$, $t_0 \sim 70{,}000$.

Of course, haven't really examined enough models to observe saturation reliably...

Thus, just as for gauge groups, there is a tremendous degeneracy in the space of heterotic string models, with many distinct models with different gauge groups and particle content sharing exactly the same value of Λ . There are flat "directions" even in the non-SUSY landscape.

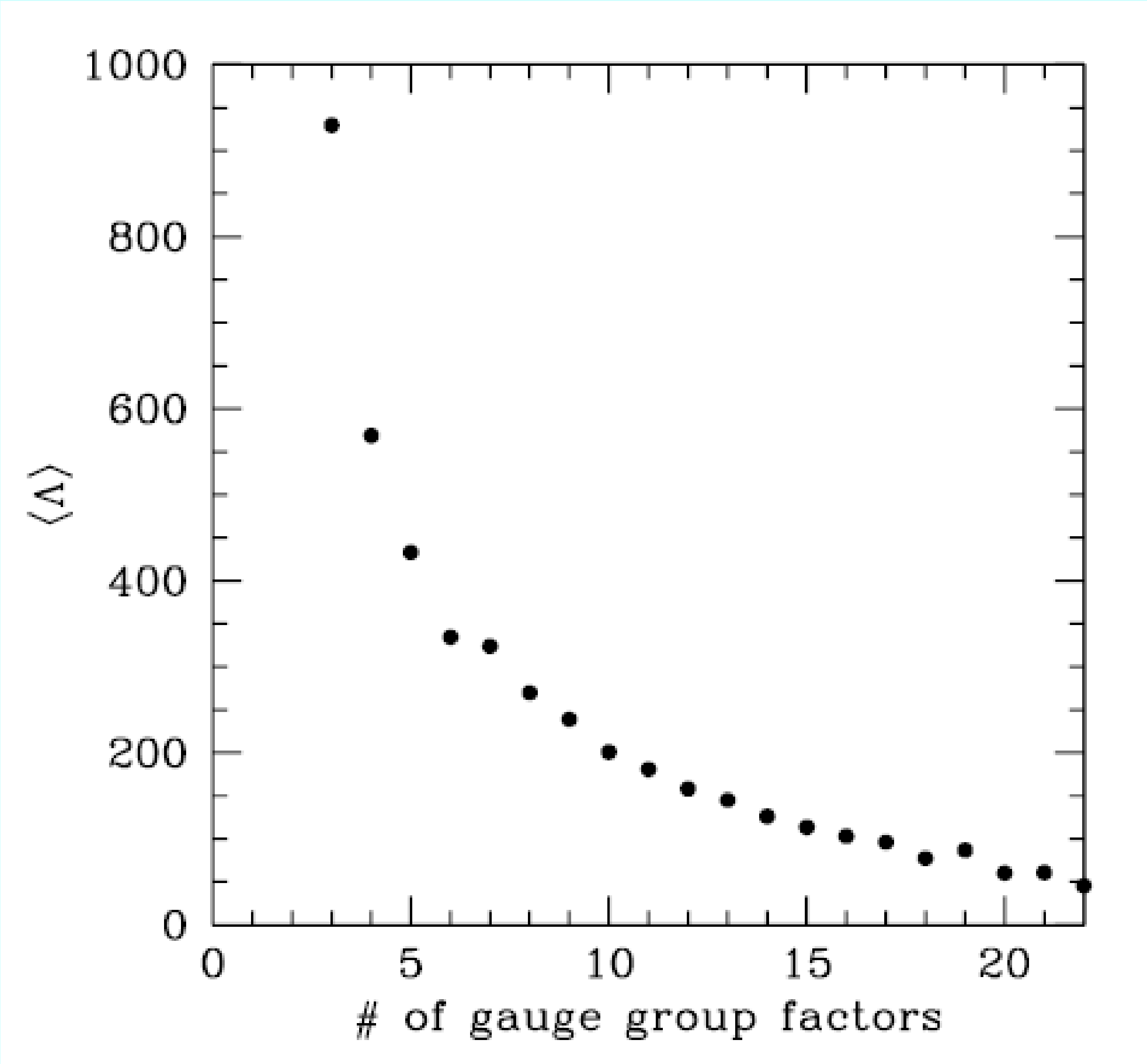
Stringy symmetries across infinite towers of massless and massive string states matter!

Having more string models does *not* necessarily imply more values of Λ !

Does this persist beyond one-loop? Expectation YES...

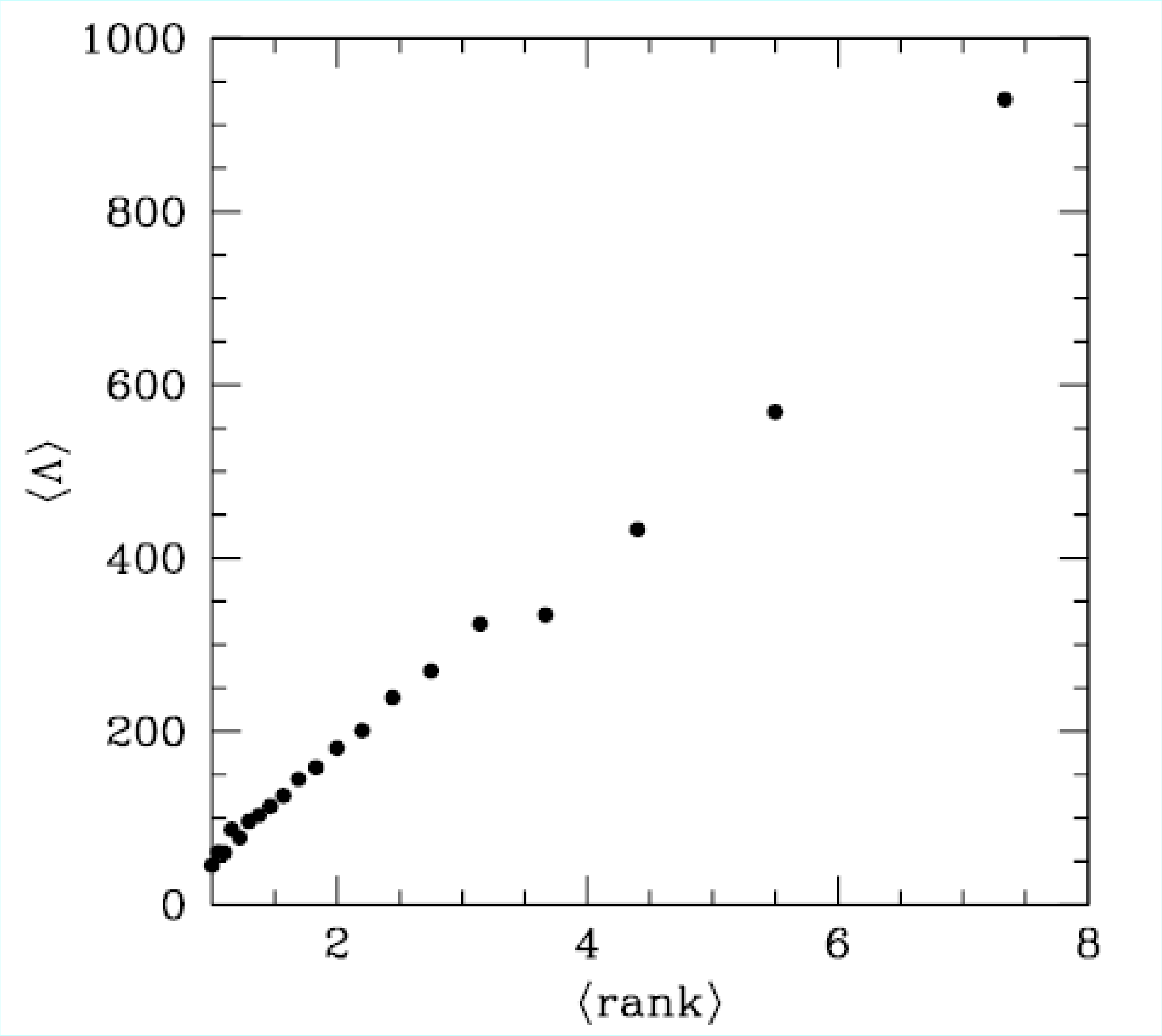
Are there significant correlations between gauge groups and Λ?

Yes! Look at Λ versus degree of shatter:



- These are *statistical averages* across all models with same degree of shatter.
- More twists tends to lead to smaller one-loop vacuum amplitudes.

Plot the same data versus average rank of factors = 22/f:



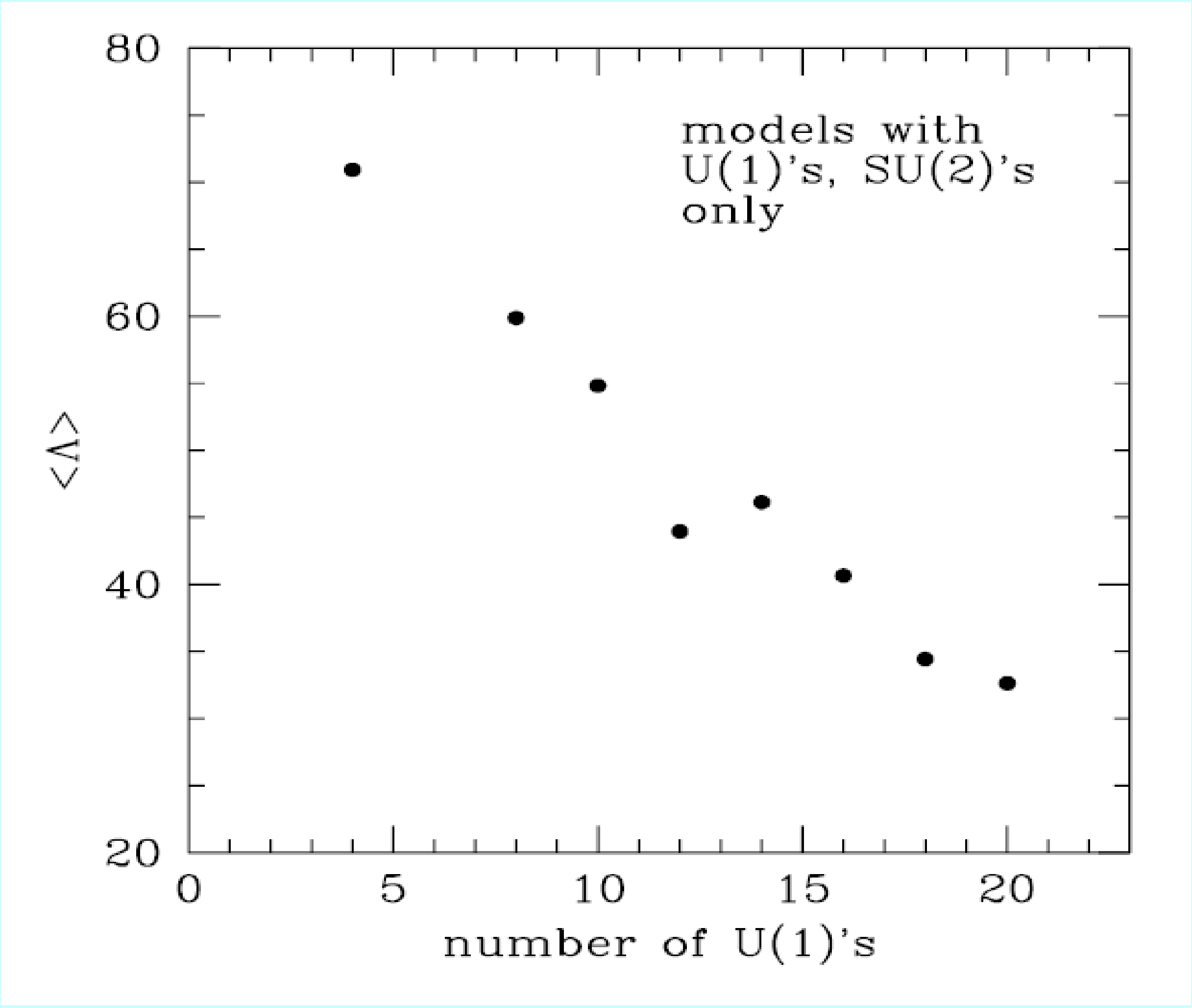
- Statistically almost a *linear* relationship.
- Suggests that contributions from vector representations dominate, with scalars cancelling against spinors and other higher reps.

Big groups lead to big Λ .

Go one step further, hold <rank> fixed.

Is Λ correlated with "abelianity" --- abelian vs. non-abelian gauge groups?

Look at string models with U(1)'s and SU(2)'s only (completely shattered). Find...



More SU(2)'s.
Very non-abelian.

More U(1)'s.
Very abelian.

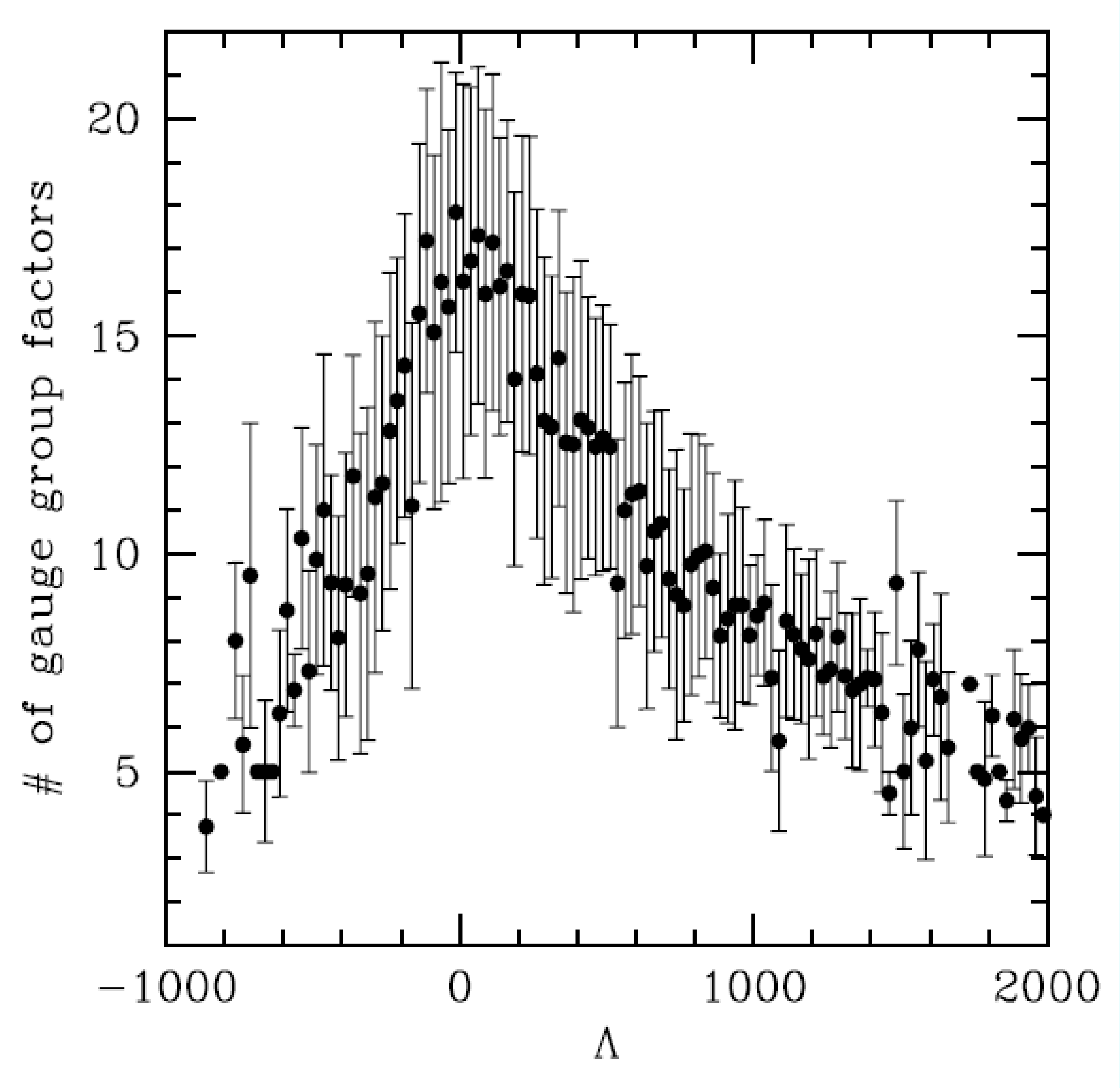
Another strong correlation, independent of previous one.

How does Λ depend on *cross-correlations* of gauge-group factors?

	U_1	SU_2	SU_3	SU_4	SU_5	$SU_{>5}$	SO_8	SO_{10}	$SO_{>10}$	$E_{6,7,8}$	SM	PS
U_1	104.6	104.6	83.2	112.9	110.7	162.3	131.2	172.1	238.8	342.2	80.8	107.6
SU_2		120.7	80.8	109.1	106.6	157.1	155.5	167.9	282.6	442.5	80.4	103.9
SU_3			85.9	90.9	113.3	136.1	117.6	162.8	193.5	220.2	83.0	88.3
SU_4				115.2	115.0	150.9	129.1	166.7	235.3	314.2	88.9	110.5
SU_5					135.9	156.3	128.1	191.6	199.2	—	107.7	110.3
$SU_{>5}$						200.9	156.4	203.2	274.5	370.7	133.5	142.8
SO_8							192.7	167.5	301.6	442.8	115.3	123.3
$S0_{10}$								207.8	253.4	289.3	166.0	159.6
$SO_{>10}$									417.4	582.8	190.8	220.0
$E_{6,7,8}$										1165.9	220.2	272.3
SM											82.5	85.5
PS												104.9
total:	108.8	121.4	83.2	113.8	110.7	162.2	163.0	173.0	298.5	440.2	80.8	108.3

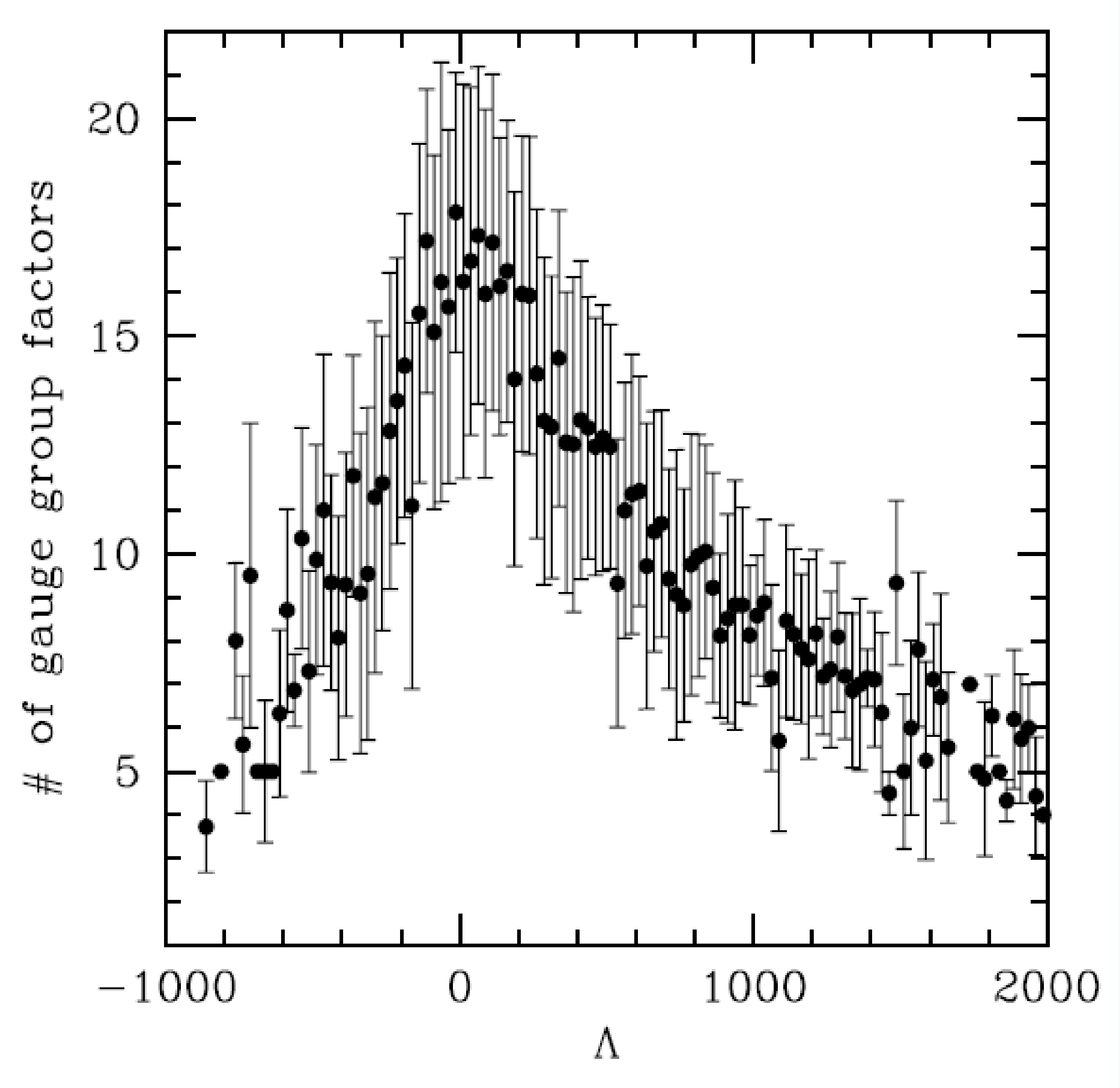
- Table shows average Λ for models with corresponding gauge group combinations.
- Models with SM gauge group have smallest values of Λ .

One can also consider the *inverse:*
For each Λ bin, calculate average number of gauge group factors for models in that bin...



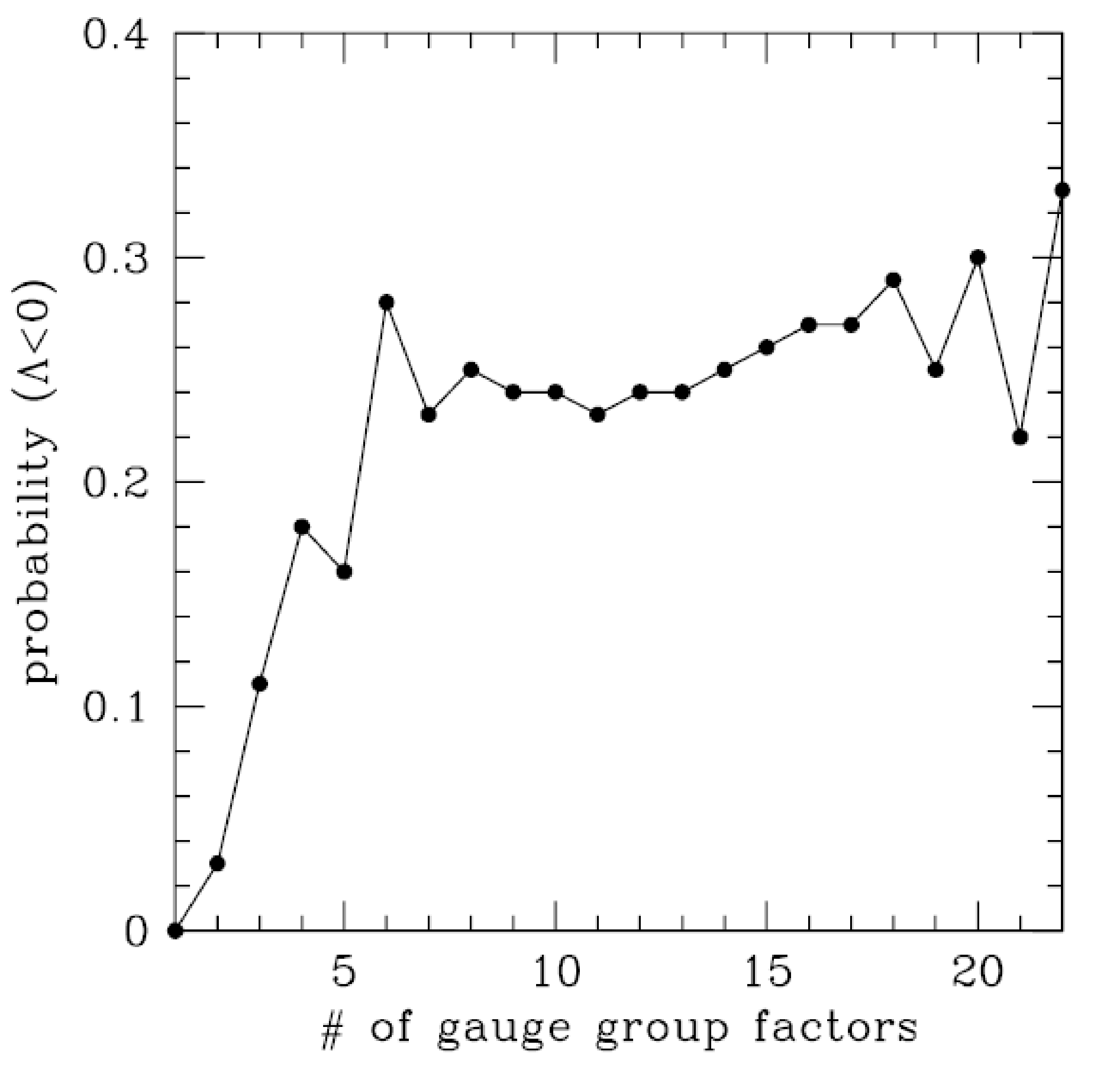
- Choosing Λ restricts number of factors to fairly narrow range (not true for inverse correlation!)
- Models with small $|\Lambda|$ have completely shattered gauge groups.

One can also consider the *inverse:*
For each Λ bin, calculate average number of gauge group factors for models in that bin...



- Fundamental limit on shattering ==> minimum size of Λ ?
- Conversely, if model with acceptable Λ is found, overwhelmingly more likely to have SM than any GUT extension.

What is the probability that a randomly chosen heterotic string model has a negative Λ (i.e., positive λ)?



- No significant probability until shatter reaches 4-5.
- Probability then remains constant as further shattering occurs.
- Overall probability averages to 27%.

Now let's examine how all of this depends on the presence or absence of SUSY.

To do this, we enlarge our data set of heterotic models to include:

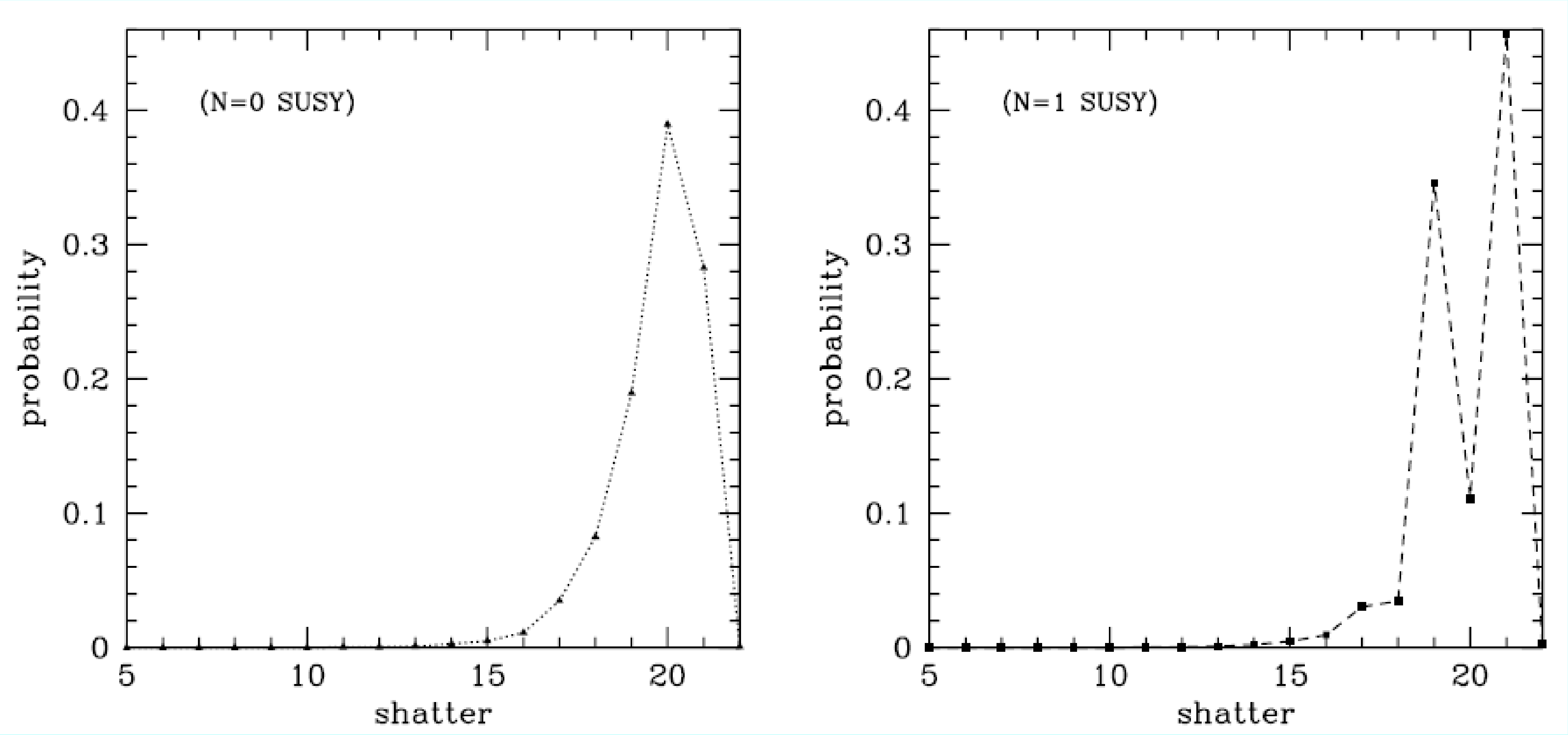
- 4.95×10^6 models without SUSY, tachyon-free
- 3.77×10^6 models with $N=1$ SUSY
- 0.49×10^6 models with $N=2$ SUSY
- 1106 models with $N=4$ SUSY

(largest sets of heterotic string models ever constructed...)

How do our previous distributions depend on N ?

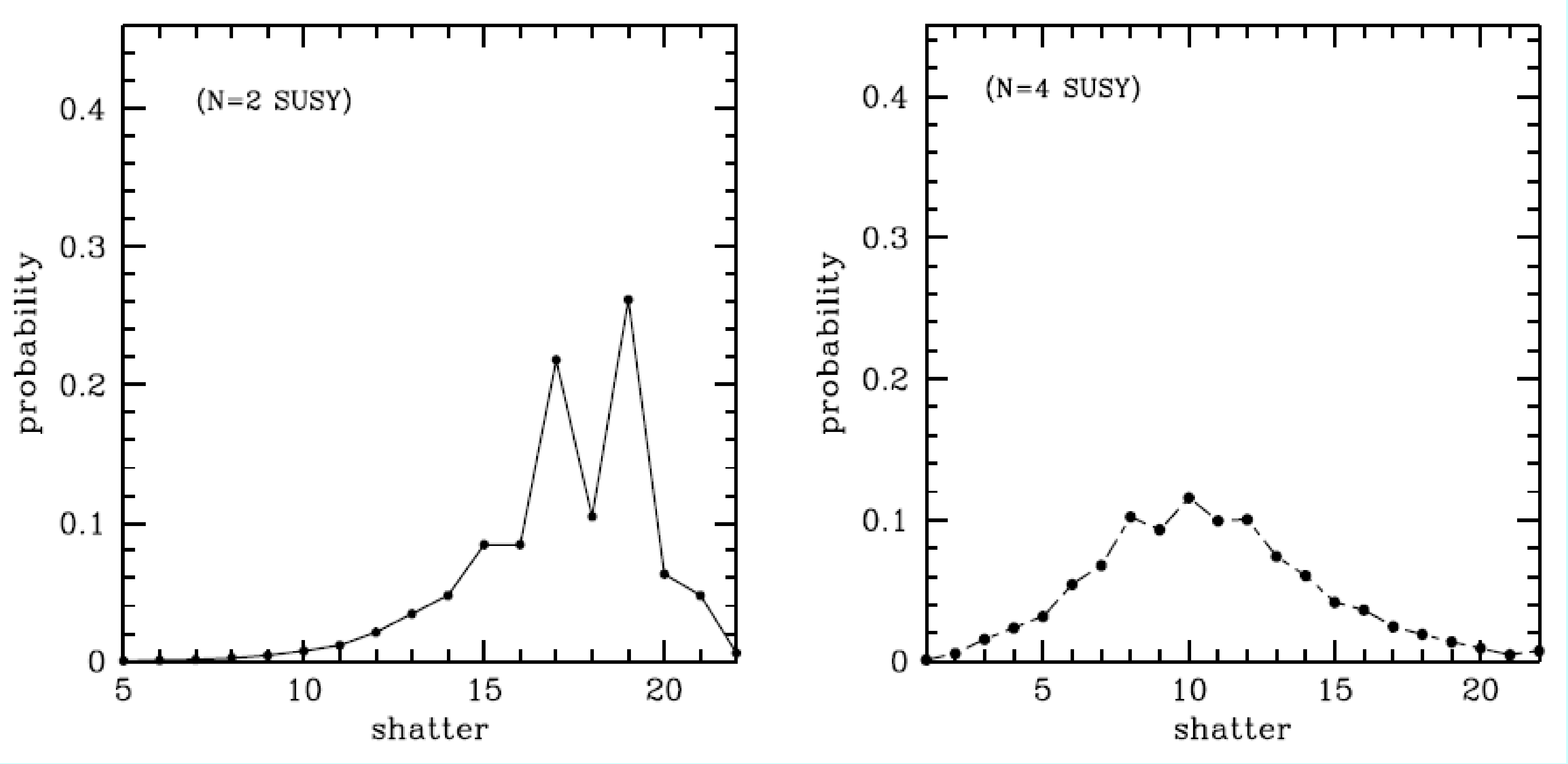
- KRD, M. Lennek, D. Senechal, V. Wasnik, arXiv:0704.1320

Which degrees of shatter are most likely?



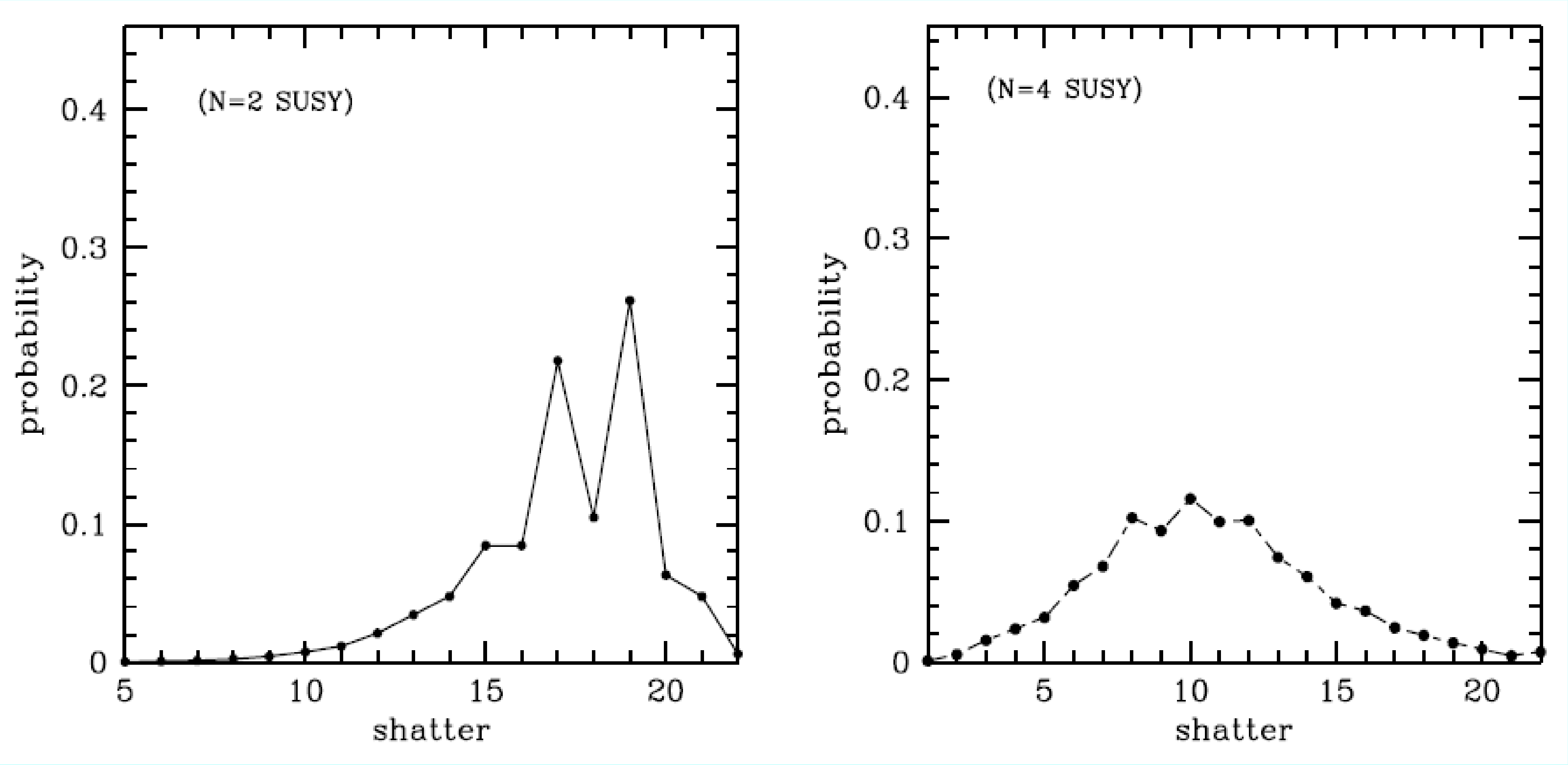
- In both cases: large shatters dominate…

But as we increase the amount of supersymmetry...



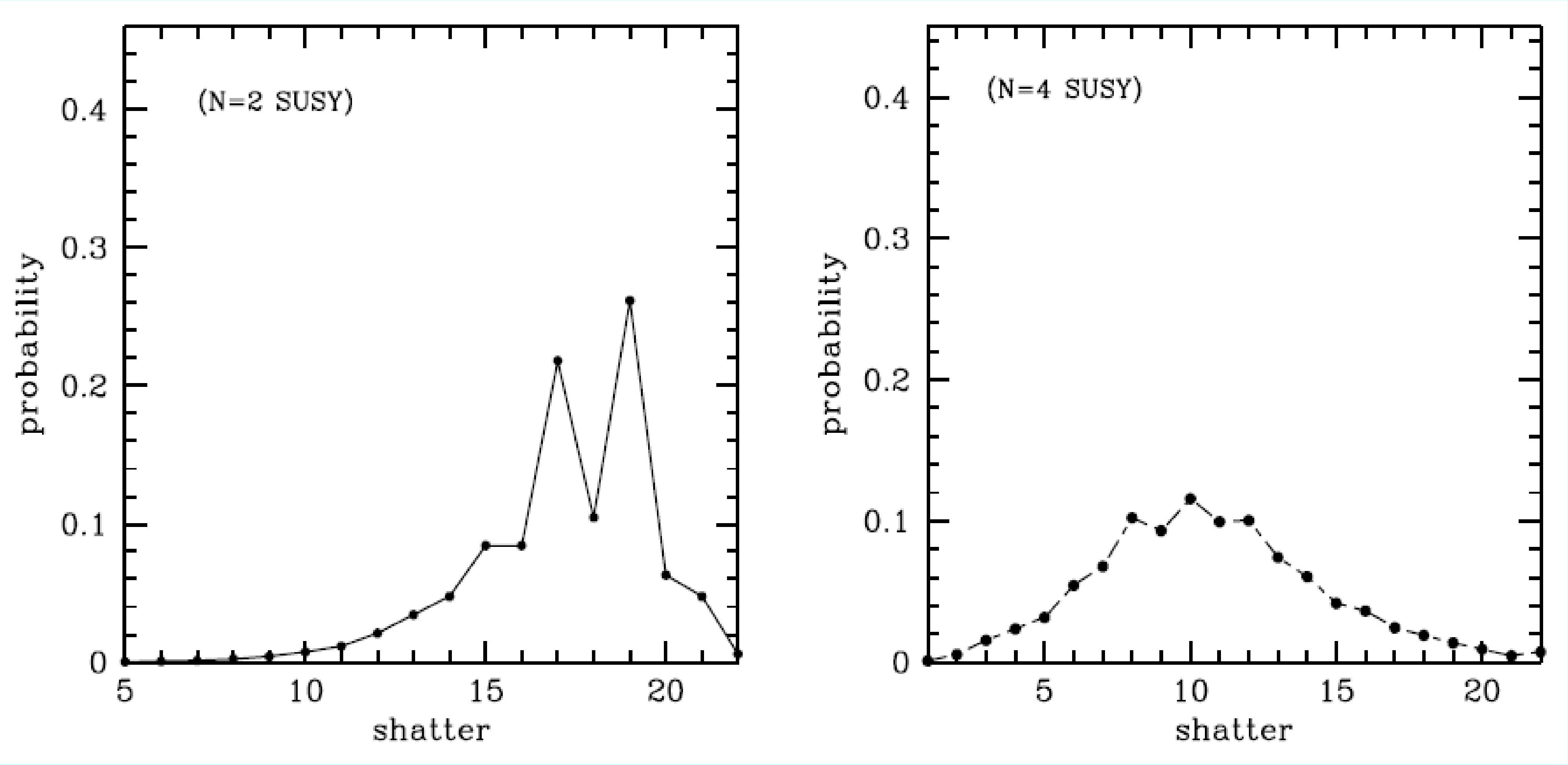
... the peak probability broadens significantly and shifts somewhat towards smaller shatters!

But as we increase the amount of supersymmetry...



Thus increasing SUSY tends to favor models with *larger* gauge-group factors.

But as we increase the amount of supersymmetry...



Interesting observation: N=4 shatter plot has an approximate *reflection* symmetry around shatter = 10. Exact? General property of maximal SUSY landscape? Prove analytically?

Ordinarily, we could now proceed to examine other important features of our data set…

- Relative populations of models with different degrees of SUSY across the landscape…
- Relative probabilities of different gauge groups as functions of the numbers of supersymmetries…
- etc…

But before we can do this, we must first deal with an important computational issue.

This is a subtle issue which is generic
to statistical landscape studies of this type:

The problem of floating correlations

This problem has not been discussed previously in the literature, but it turns out to play a huge role in obtaining meaningful statistical results from a data set to which one has only limited computational access.

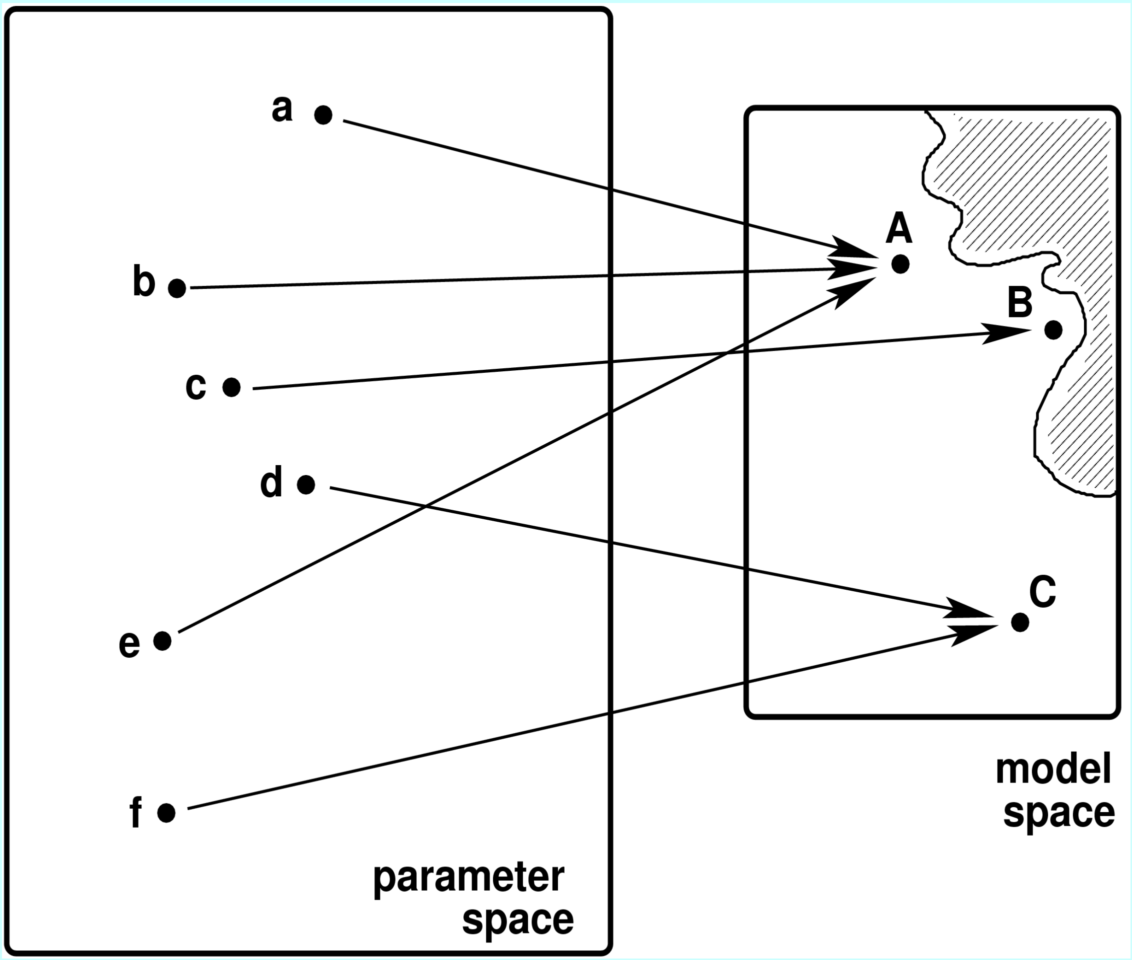
- KRD and M. Lennek, hep-th/0610319 (PRD)

The problem of floating correlations is the observation that some statistical correlations are *unstable* --- they "*float*" (or evolve) as the sample size increases.

Why does this happen?

Essentially, as we continue to randomly generate models, it gets harder and harder to find new (i.e., distinct) models. Thus, physical characteristics which were originally "rare" are often forced to become less "rare" as the sample size increases and we probe more deeply into the space of models.

In particular,
consider the process of randomly generating string models…



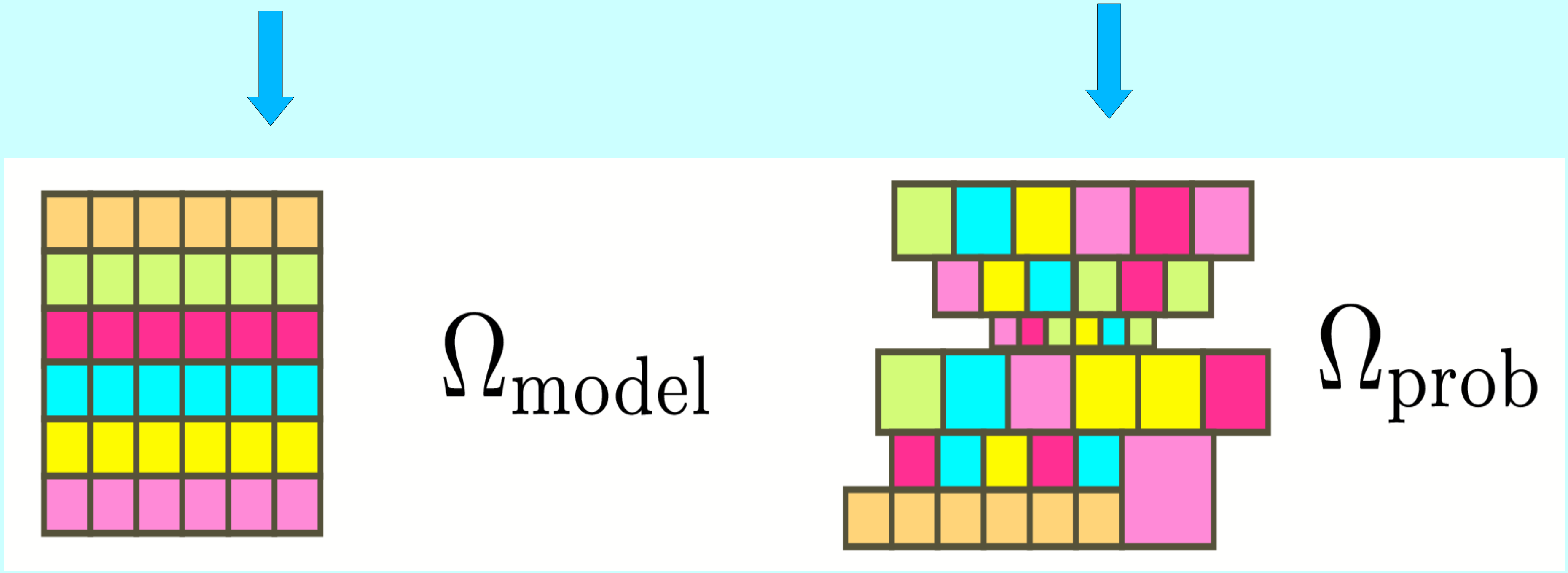
- One must generically employ a model-construction technique which specifies models according to some set of internal parameters (e.g., fluxes, orbifold twists, boundary conditions or phases, Wilson lines, etc.)

- Each set of parameters maps to a single model, but the mapping is rarely unique!

Thus some models are much more likely to be generated than others! This feature is essentially *unavoidable*.

Thus, we don't see the model space directly:

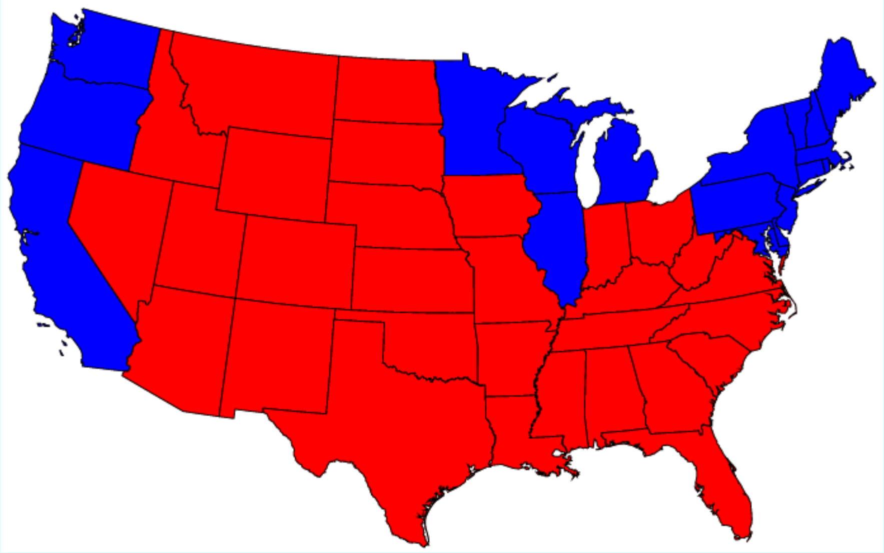
We see a deformed version of it, a "probability space":



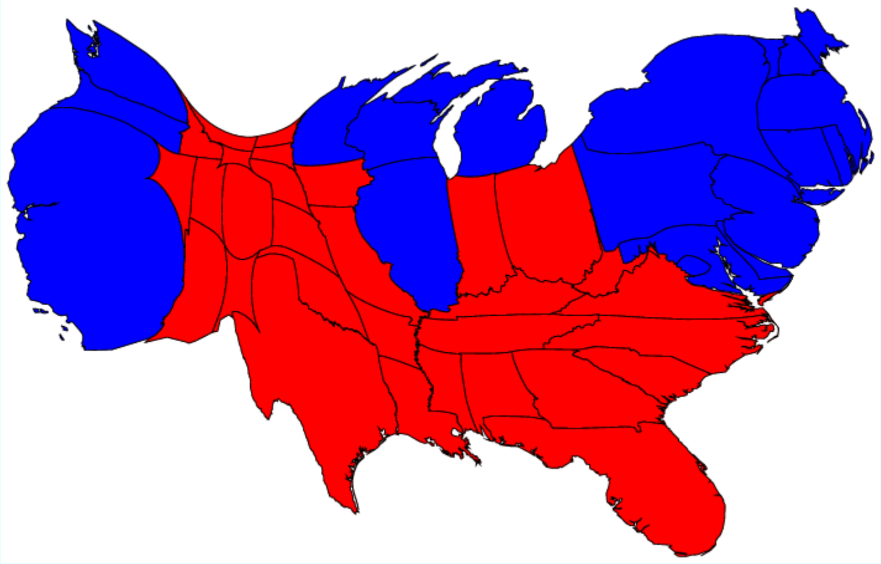
Does this difference matter for our statistical correlations between physical observables?

Yes, if the physical properties are somehow correlated with these probability deformations.

To use a real-world example, it's the difference between this:

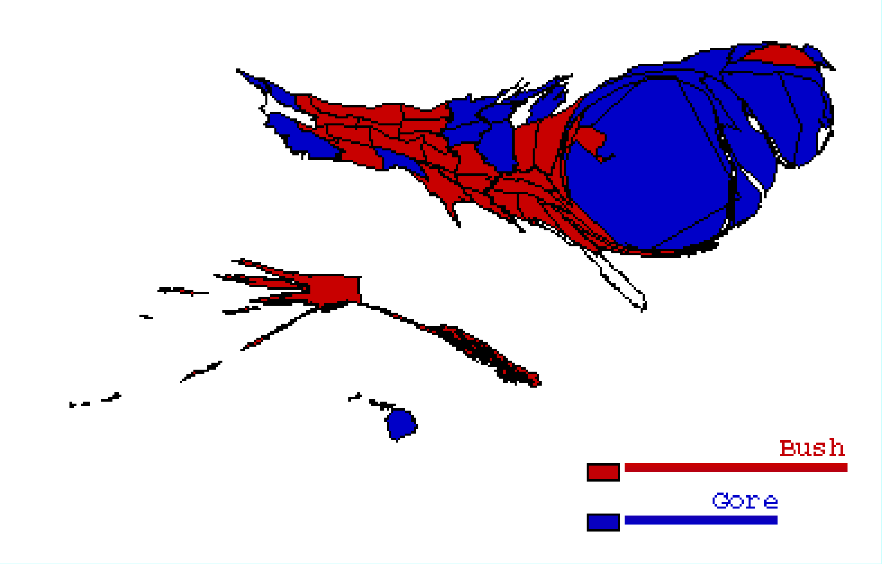


and this:



Cartogram based on population.

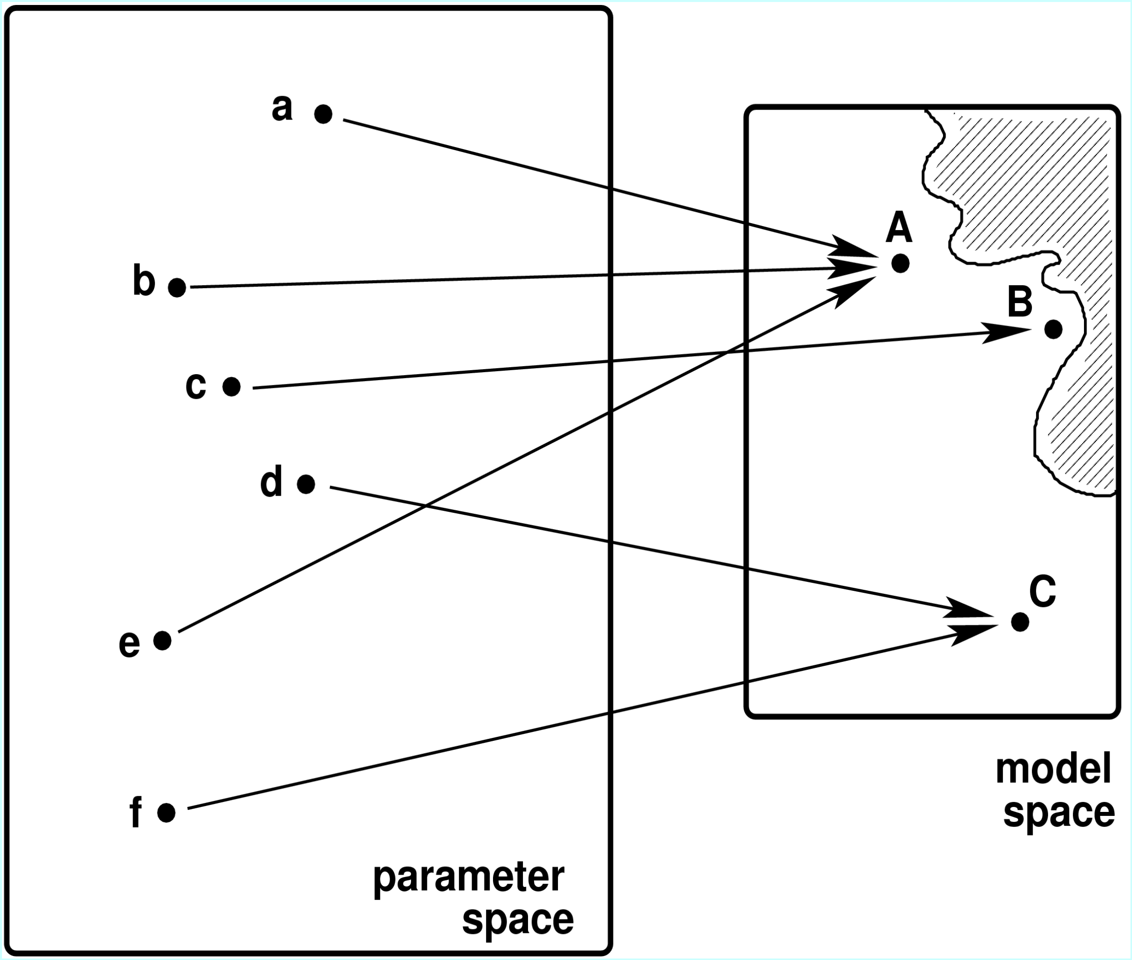
...or even this:



Cartogram based on population density.

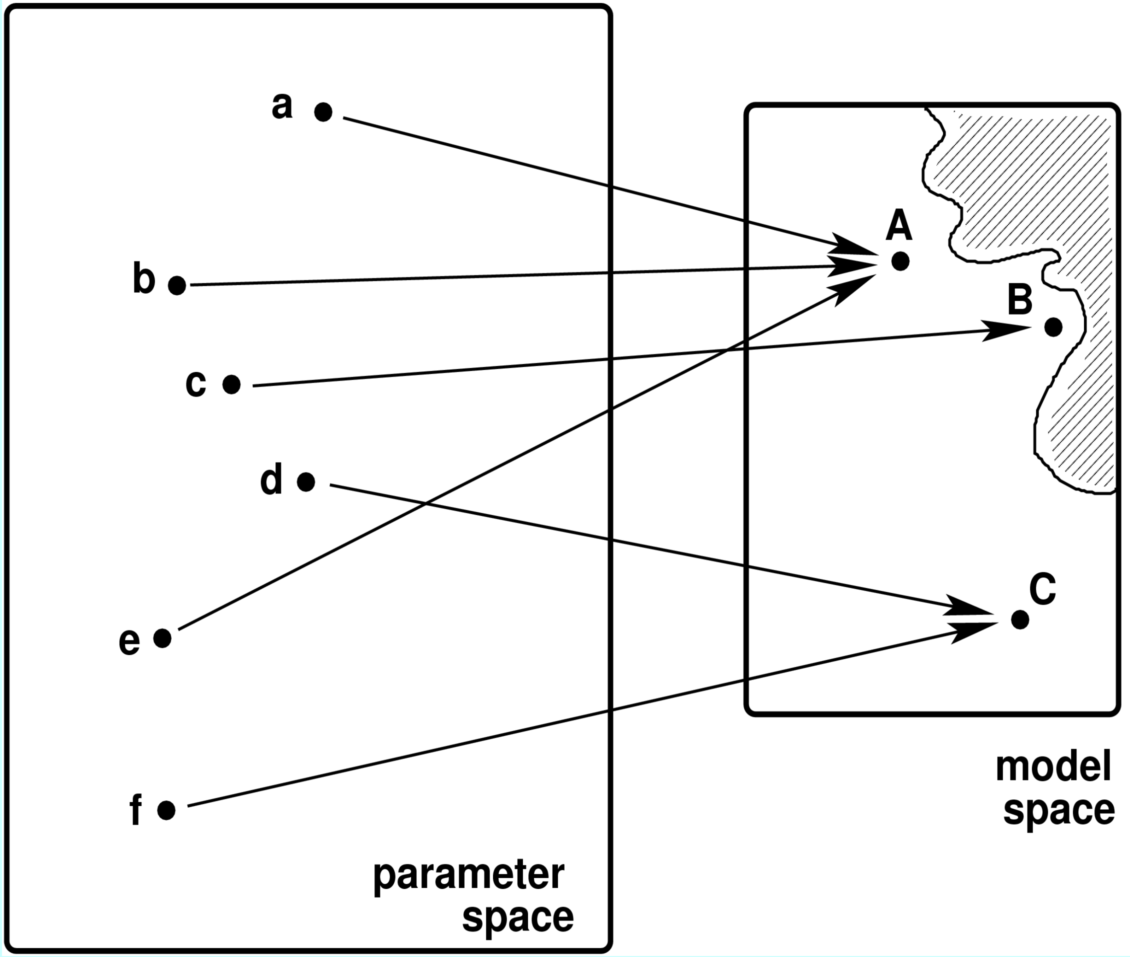
Sadly, these things *do* matter and can affect outcomes.

How can we get around this problem?



- Partial solution: don't count the “new” model if it's already in the data set. Consider it a “failed attempt”, disregard this case, and try again.

How can we get around this problem?



- But we are still not finding the very “rare” models (such as Model B), close to the “unreachable” region. It will take a considerably larger data set before we will stumble across such rare models, and we have *no information* about where they are, how common they are, or whether they even exist!

This is the whole problem: we do not have computational access to the entire landscape! Thus, our statistical data “floats” as we keep digging for new nuggets (which, since they are still “new”, are necessarily “rare”).

What we need is a way of extracting information (even if only limited information) about the full landscape on the basis of only partial information.

Analogous to lattice gauge theory: need to extract information about the continuum limit on the basis of calculations done at finite lattice spacing.

Solution:

- Restrict attention to relative *ratios* of probabilities of models with different characteristics.
- But calculate these ratios only when the spaces of models with these characteristics *are equally explored.*

Of course, we need a measure for "equally explored". How can we judge how deeply we have penetrated into a particular model space?

Solution: Look at *number of attempts to generate a model with a specified characteristic*. If it is easy to generate new models of a given type, then the corresponding space of models of that type is relatively unexplored. As we progress, it gets much harder to find new models of that type and the number of failed attempts per new model increases.

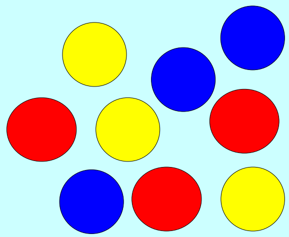
Thus, by measuring numbers of models found against numbers of *attempts* to generate new models, and comparing this ratio for two different groups of models, we can extract information about the relative volumes of their corresponding full model spaces and thereby deduce their true relative probabilities.

This assumes, of course, that the physical properties of interest are correlated with the probability deformations of their corresponding model spaces (so that “rare” models tend to have Property #1, “common” models have Property #2, etc.).

This will be the case if the biases associated with the underlying model-construction technique are correlated with the physical properties of the models they produce.

We have found this to be true in most cases.

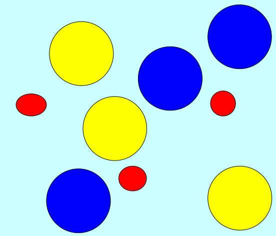
Example: Plucking balls from an urn.



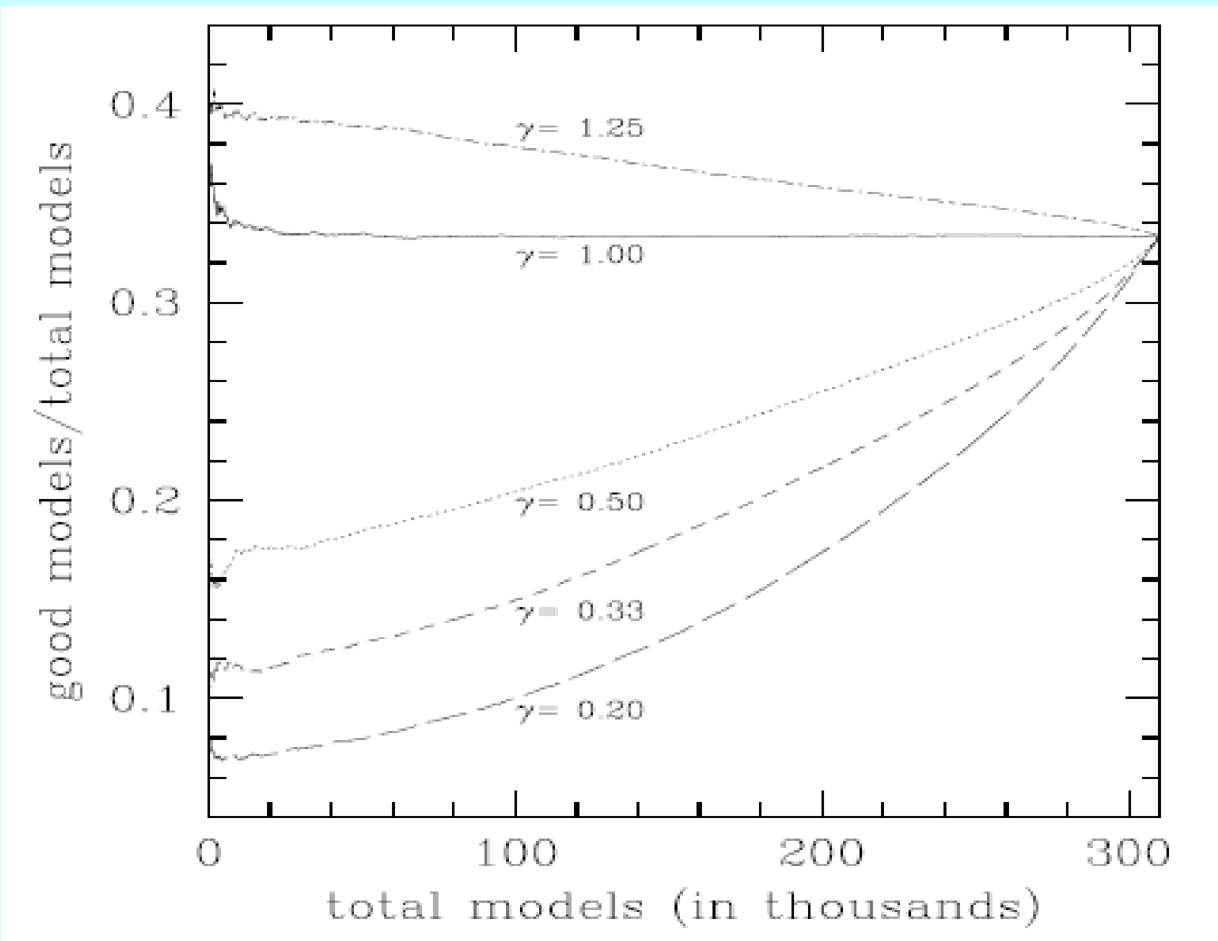
An urn contains 300,000 balls of different colors. One third of the balls are red. We seek to know what fraction of balls in the urn are red, and we try to determine this by choosing a ball randomly from the urn, noting its color, marking it for future identification, replacing the ball in the urn, mixing, and then repeating over and over.

If all balls are treated equally (no bias), approximately one third of all balls selected will be red. This will not vary significantly with sample size.

However, suppose the red balls have a different size than the others, so that the probability of picking a red ball from the urn on a given try is γ times the probability of picking a ball of any other color.



What fraction of selected balls will be red?
Clearly this "floats" with the sample size:



True fraction emerges only upon full exploration of the urn.

But suppose we don't have enough time/ability to wait that long and we don't know γ. What can we do?

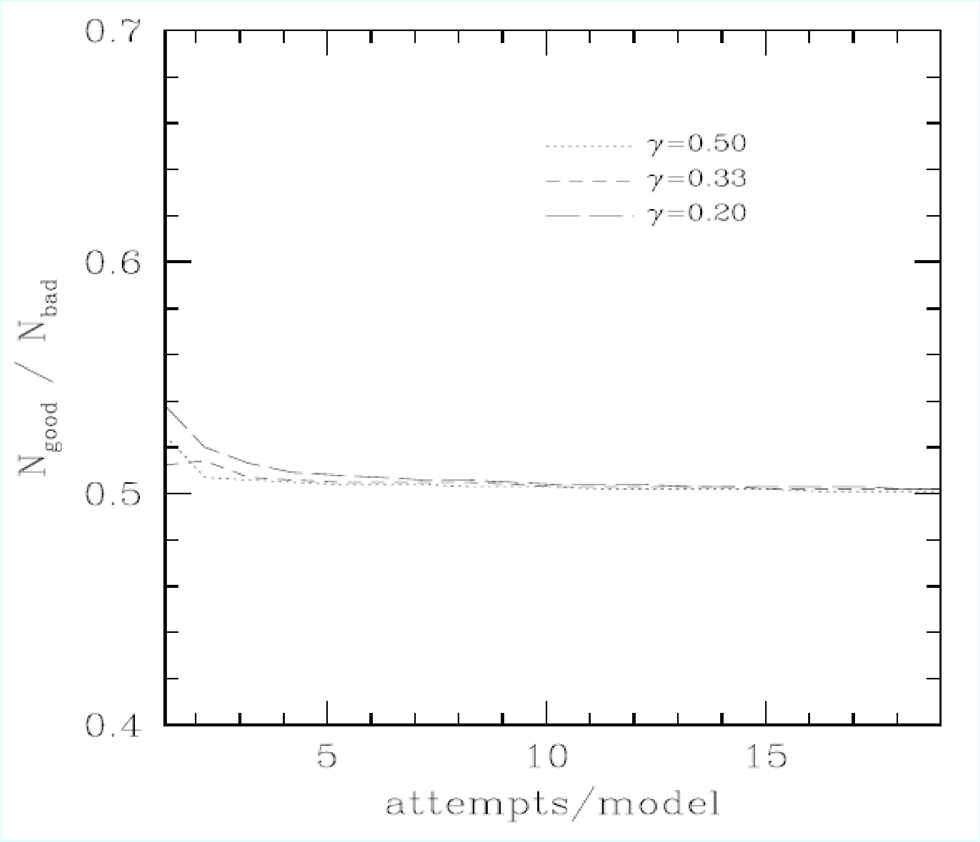
Keep a running record of

- X_{red} = number of failed "red" attempts to find the last new red ball
- X_{other} = number of failed "other" attempts to find a new ball of any other color.

Then

$$\frac{\text{Number of red balls in urn}}{\text{Number of other balls in urn}} = \frac{\text{\# red balls that have been found}}{\text{\# other balls that have been found}}$$

evaluated at values for which $X_{red} = X_{other}$!!



"Continuum" limit reached quite quickly regardless of chosen X!

In fact, the true computational situation we face for the landscape is even more complicated ---

- There can be a whole *spectrum of different sizes* (intrinsic probabilities) for the different balls (string models).
- There is no guarantee that the *sizes* (intrinsic probabilities) of the balls (models) are in any way correlated with their *colors* (physical characteristics).

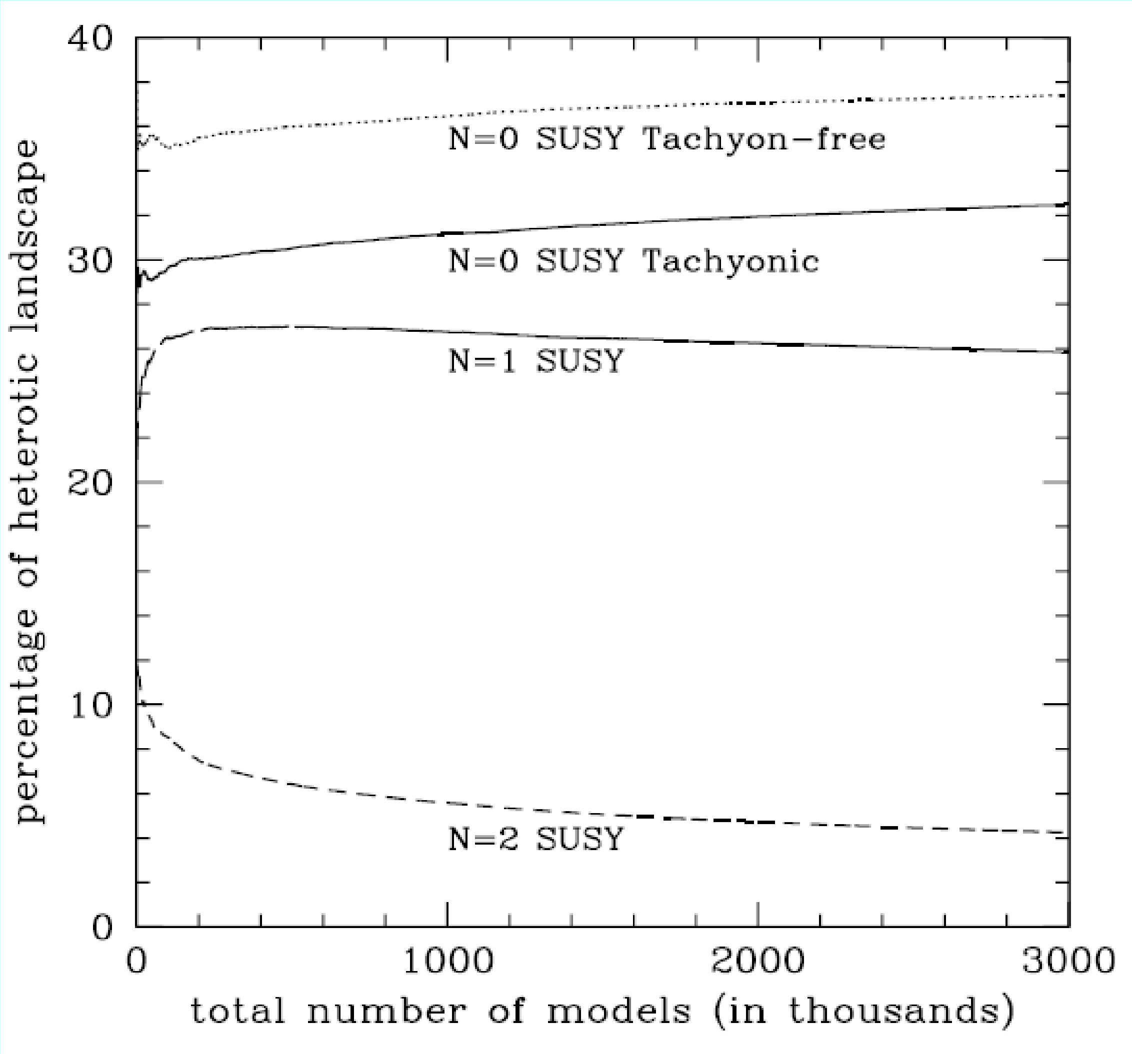
➡ In general, there can be a huge "CKM matrix" between colors and sizes, all of whose entries are essentially unknown!

Need methods of extracting meaningful statistical information, even for such general situations.

- KRD and Lennek, hep-th/0610319 (PRD)

Using these techniques, for example, we can estimate the relative probabilities of models with different levels of SUSY...

First, note that we cannot just take a census of the models we have randomly generated --- these ratios *float*!



- As we generate more and more models, it becomes increasingly likely to generate non-SUSY models than SUSY models.
- Suggests that at any given moment, we have already explored more of the SUSY model space than the non-SUSY model space.

Indeed, looking at attempts/model, we verify this...

SUSY class	# distinct models	# attempts	avg. attempts/model
$\mathcal{N}=0$ (tachyonic)	1 279 484	3 810 838	2.98
$\mathcal{N}=0$ (tachyon-free)	4 946 388	18 000 000	3.64
$\mathcal{N}=1$	3 772 679	24 200 097	6.41
$\mathcal{N}=2$	492 790	13 998 843	28.41
$\mathcal{N}=4$	1106	6 523 277	5 898.08
Total:	10 492 447	66 533 055	6.34

- We have penetrated the SUSY model spaces much more fully than the non-SUSY model space.
- Thus, as we continue to generate new models, it becomes harder and harder to find new SUSY models compared with non-SUSY models.
- Relative ratios of SUSY and non-SUSY models therefore *float* as a function of sample size, as already seen.

Indeed, looking at attempts/model, we verify this...

SUSY class	# distinct models	# attempts	avg. attempts/model
$\mathcal{N}=0$ (tachyonic)	1 279 484	3 810 838	2.98
$\mathcal{N}=0$ (tachyon-free)	4 946 388	18 000 000	3.64
$\mathcal{N}=1$	3 772 679	24 200 097	6.41
$\mathcal{N}=2$	492 790	13 998 843	28.41
$\mathcal{N}=4$	1106	6 523 277	5 898.08
Total:	10 492 447	66 533 055	6.34

Thus, we must compare relative numbers of models in each SUSY class only at those points during our search at which their average attempts/model are equal!

It is only *these* ratios which will be truly stable, i.e.,

- independent of sample size
- independent of the chosen reference value of attempts/model.

In this way, we extract the "continuum" limit, ultimately obtaining...

SUSY class	% of heterotic landscape
$\mathcal{N}=0$ (tachyonic)	32.1
$\mathcal{N}=0$ (tachyon-free)	46.5
$\mathcal{N}=1$	20.9
$\mathcal{N}=2$	0.5
$\mathcal{N}=4$	0.003

- Nearly half of the heterotic landscape is non-SUSY but tachyon-free!
- The SUSY portion of the heterotic landscape represents less than ¼ of the full landscape, even at the string scale!
- Models exhibiting extended (N>1) SUSY are exceedingly rare, representing less than 1% of the full landscape.

In this way, we extract the “continuum” limit, ultimately obtaining...

SUSY class	% of heterotic landscape
$\mathcal{N}=0$ (tachyonic)	32.1
$\mathcal{N}=0$ (tachyon-free)	46.5
$\mathcal{N}=1$	20.9
$\mathcal{N}=2$	0.5
$\mathcal{N}=4$	0.003

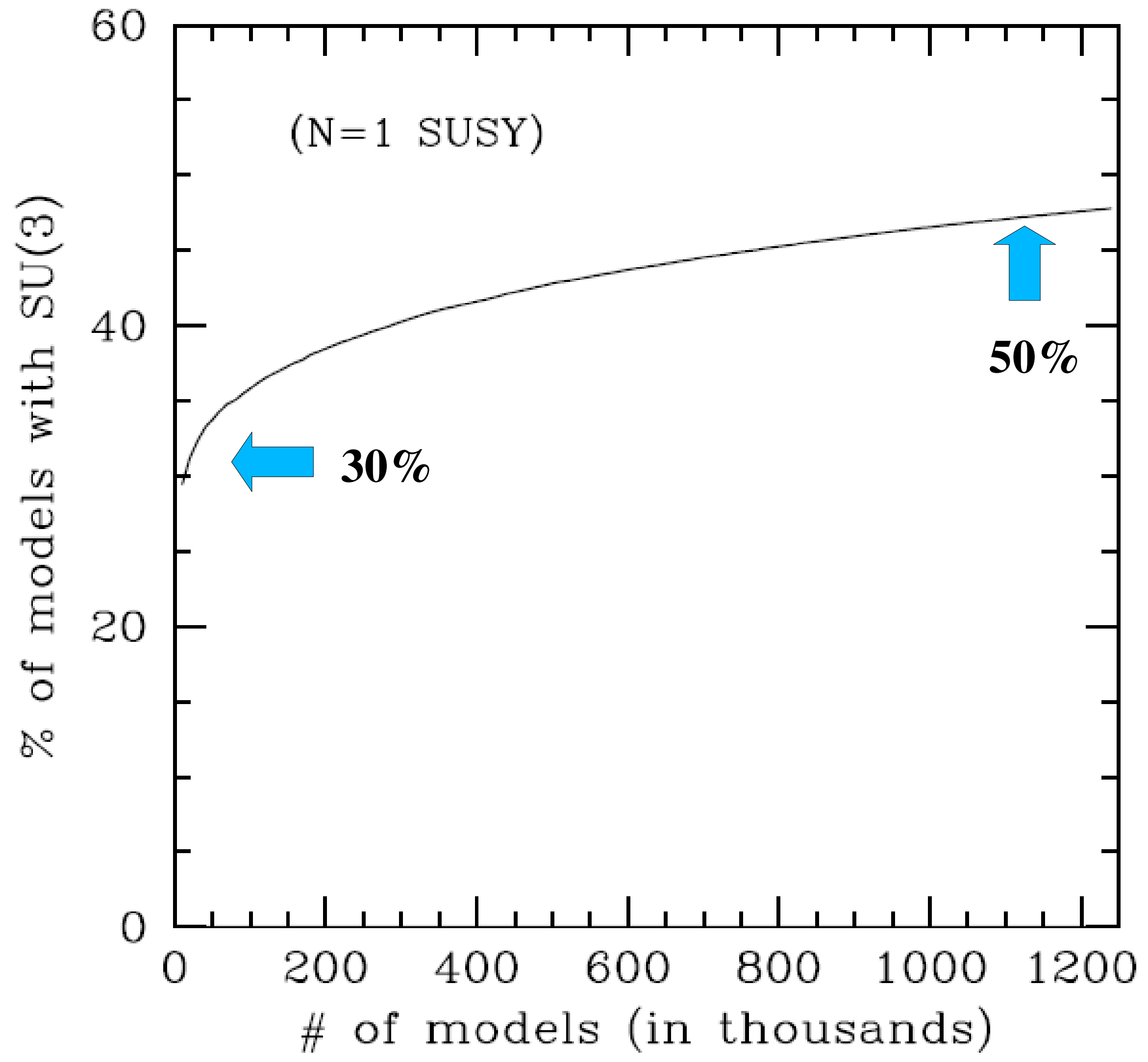
Will these results change after moduli stabilization?

Seems unlikely, since most modern methods of moduli stabilization (fluxes, superpotentials, etc.) tend to further break (rather than restore) spacetime SUSY.

Such results therefore tend to shift burden of proof onto SUSY enthusiasts ... a dramatic reframing of the underlying question!

Another example: gauge-group populations versus SUSY.
These ratios also float significantly.

e.g., probability of realizing SU(3) gauge group in N=1 models...



- SU(3) *seems* rare for small sample sizes, but becomes significantly less rare as sample size grows!
- ***How high does it get?*** Need to extract continuum limit relative to other gauge groups, eventually on an absolute scale.
- Using previous methods, can ultimately extract stable results...

How likely are different gauge group factors?

gauge group	$\mathcal{N}=0$	$\mathcal{N}=1$	$\mathcal{N}=2$	$\mathcal{N}=4$
U_1	99.9	94.5	68.4	89.6
SU_2	62.46	97.4	64.3	60.9
SU_3	99.3	98.0	93.0	45.1
SU_4	14.46	30.0	39.0	53.5
SU_5	16.78	43.5	66.3	33.8
$SU_{>5}$	0.185	1.7	10.6	73.0
SO_8	0.482	1.6	6.2	21.1
SO_{10}	0.084	0.2	1.6	18.7
$SO_{>10}$	0.005	0.038	0.77	7.5
$E_{6,7,8}$	0.0003	0.03	0.16	11.5

- SU(3) has "floated" all the way to 98%.
- SU(n+1) groups preferred over SO(2n) groups for each rank n.

How likely are different gauge group factors?

gauge group	$\mathcal{N}=0$	$\mathcal{N}=1$	$\mathcal{N}=2$	$\mathcal{N}=4$
U_1	99.9	94.5	68.4	89.6
SU_2	62.46	97.4	64.3	60.9
SU_3	99.3	98.0	93.0	45.1
SU_4	14.46	30.0	39.0	53.5
SU_5	16.78	43.5	66.3	33.8
$SU_{>5}$	0.185	1.7	10.6	73.0
SO_8	0.482	1.6	6.2	21.1
SO_{10}	0.084	0.2	1.6	18.7
$SO_{>10}$	0.005	0.038	0.77	7.5
$E_{6,7,8}$	0.0003	0.03	0.16	11.5

- Groups with smaller ranks are much more common than groups with larger ranks... true for all levels of SUSY.

How likely are different gauge group factors?

gauge group	$\mathcal{N}=0$	$\mathcal{N}=1$	$\mathcal{N}=2$	$\mathcal{N}=4$
U_1	99.9	94.5	68.4	89.6
SU_2	62.46	97.4	64.3	60.9
SU_3	99.3	98.0	93.0	45.1
SU_4	14.46	30.0	39.0	53.5
SU_5	16.78	43.5	66.3	33.8
$SU_{>5}$	0.185	1.7	10.6	73.0
SO_8	0.482	1.6	6.2	21.1
SO_{10}	0.084	0.2	1.6	18.7
$SO_{>10}$	0.005	0.038	0.77	7.5
$E_{6,7,8}$	0.0003	0.03	0.16	11.5

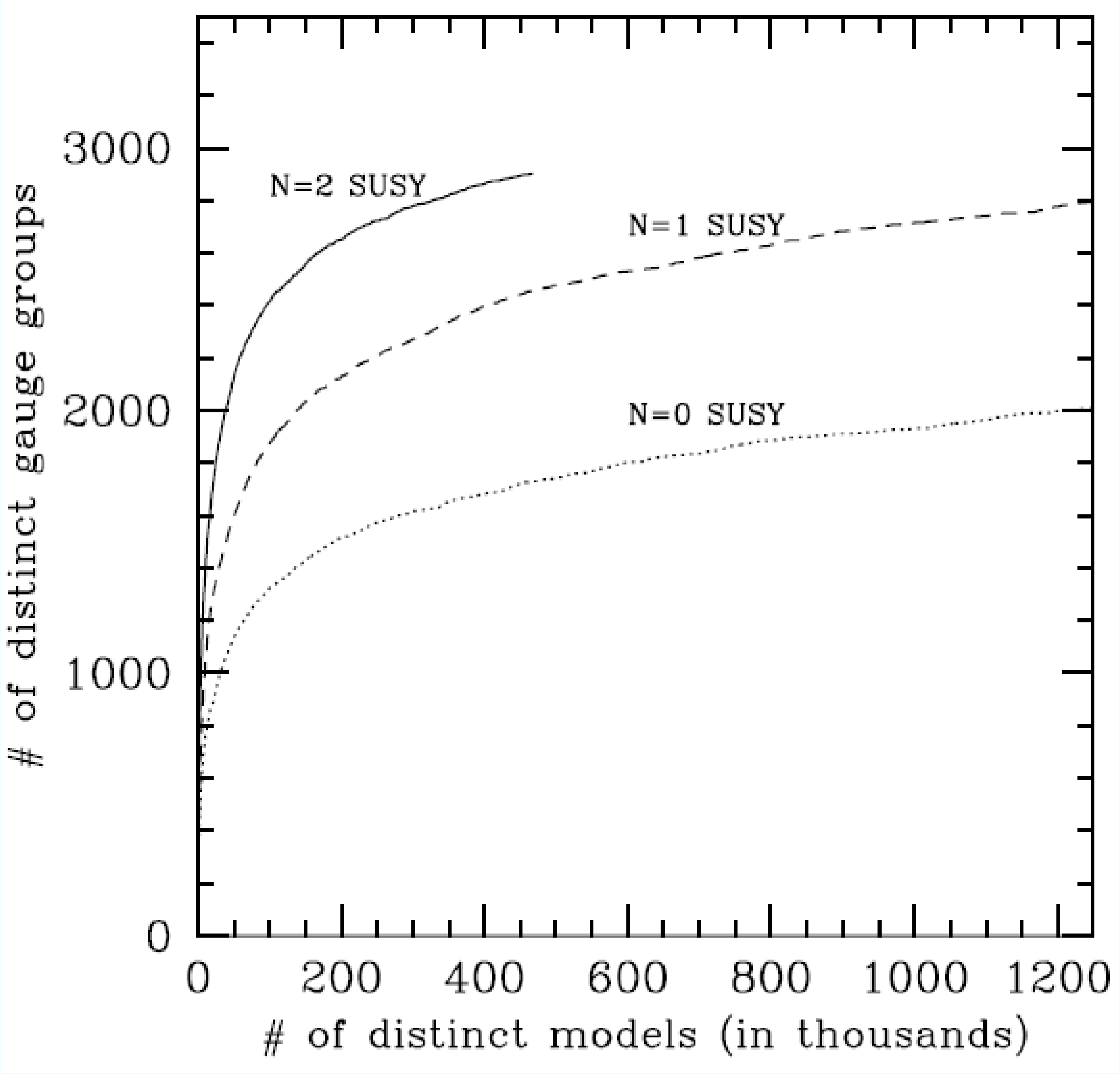
- Finally, gauge-group factors comprising the SM are much more common than any of its grand-unified extensions.

Or, collapsing these results in terms of the entire landscape versus its supersymmetric subset...

gauge group	entire landscape	SUSY subset
U_1	98.00	93.89
SU_2	73.22	96.62
SU_3	98.85	97.88
SU_4	19.42	30.21
SU_5	25.37	44.03
$SU_{>5}$	0.73	1.92
SO_8	0.87	1.71
SO_{10}	0.13	0.23
$SO_{>10}$	0.02	0.06
$E_{6,7,8}$	0.01	0.03

- Gauge groups with larger ranks are favored more strongly with SUSY than without SUSY.
- Thus, small factors such as U(1) and SU(3) are slightly *less* prevalent with SUSY... they are "sacrificed" in order to make room for larger-rank gauge groups.

Finally, how many realizable gauge groups are there as a function of the number of supersymmetries?



- In all cases, there are many more models than gauge groups.
- But as SUSY increases, string constraints get tighter, with distinct models increasingly forced to have distinct gauge groups.
- Thus, there are fewer distinct models per gauge group as SUSY increases.

Even this gauge-group multiplicity floats!

However, using our previous methods, we can extract stable ratios (relative to N=1 case) for multiplicities as well as for overall numbers of realizable gauge groups...

SUSY class	avg. multiplicity per gauge group	# of realizable gauge groups
$\mathcal{N}=0$ (tachyon-free)	1.65	1.35
$\mathcal{N}=1$	1.00	1.00
$\mathcal{N}=2$	0.89	0.03

- Evidently, requirement of avoiding tachyons in N=0 case is *less stringent* than preserving SUSY, at least as far as gauge-group multiplicities are concerned. (Opposite is true in D=10!)
- We do not quote absolute results because we could not obtain a stable overall normalization. Only relative ratios were stable.
- For N=4, each distinct string model had a unique gauge group!

And the list goes on...

- Chirality
- Numbers of fermion generations
- Hypercharge normalizations
- Gauge coupling unification
- Yukawa couplings
- String threshold corrections
- Intermediate-scale physics (SUSY-breaking, new gauge structures, ...)
- etc.

This work is in progress, due out soon.

KRD, M. Lennek, D. Senechal, V. Wasnik (to appear)

Conclusions, Prospects, and Warnings

Clearly, a statistical analysis of the string landscape has lots of potential to address questions of relevance to phenomenology --- *even without a vacuum-selection principle.*

Much more work remains to be done…

- Other phenomenological features need to be examined: particle content, etc., as already discussed.
- Develop algorithmic/statistical tools to handle analyses of this type.
- Extend analysis to broader classes of string theories (more general constructions, also non-perturbative formulations).
- Develop methods to generate large classes of *stable* vacua --- comparison of results will then indicate phenomenological role played by vacuum stability.
- Comparison with Type I results may even permit *statistical* confirmation of duality conjectures.

Indeed, the SVP will be tackling many of these questions.

But one must be aware of certain dangers...

- The "lamppost" effect --- the danger of restricting one's attention to those portions of the landscape where one has control over calculational techniques.
- The "Godel" effect --- landscape is so large that it is possible that no matter how many input "priors" one demands, there will always be another observable which cannot be uniquely predicted.
- The "bull's-eye" effect --- don't always know what the target is, since we are not certain how our low-energy world embeds into the fundamental theory (SUSY? GUTs? technicolor? something else?).

Nevertheless, despite these dangers,

- Direct examination of actual string models uncovers features and behaviors that might not otherwise be expected.
- Through direct enumeration, we gain valuable experience in the construction and analysis of phenomenologically viable string vacua.
- As string theorists, we must ultimately come to terms with the landscape. Just as in astrophysics, botany, and zoology, the first step in the analysis of a large data set is enumeration and classification.
- In cases where statistical correlations can be interpreted directly in terms of underlying physical symmetries, we have indeed extracted true predictions from the landscape.

Thus, properly interpreted, statistical landscape studies can be useful and relevant in this overall endeavor.